PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 9/04 — A1

(11) International Publication Number: WO 97/11162

(43) International Publication Date: 27 March 1997 (27.03.97)

(21) International Application Number: PCT/CA96/00605

(22) International Filing Date: 12 September 1996 (12.09.96)

(30) Priority Data:
08/532,896 22 September 1995 (22.09.95) US

(71) Applicant *(for all designated States except US)*: ENDORECHERCHE INC. [CA/CA]; 2989 de la Promenade, Ste-Foy, Québec G1W 2J5 (CA).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: LABRIE, Fernand [CA/CA]; 2989 de la Promenade, Ste-Foy, Québec G1W 2J5 (CA). LUU-THE, Van [CA/CA]; 4460 rue de l'Estuaire, Charny, Québec G6X 1C6 (CA).

(74) Agent: MITCHELL, Richard, J.; Marks & Clerk, P.O. Box 957, Station B, Ottawa, Ontario K1P 5S7 (CA).

(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: PRODUCTION AND USE OF TYPE 5 17BETA-HYDROXYSTEROID DEHYDROGENASE

(57) Abstract

A novel type 5 17β-hydroxysteroid dehydrogenase is provided. Methods of producing the enzyme and using the enzyme to identify potential compounds which inhibit or alter the activity of the enzyme are described. In addition, methods of using the gene sequence or portions thereof for probes or to produce expression-disrupting sense or antisense DNA fragments thereof, or antisense RNA, are provided.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LR	Liberia	SZ	Swaziland
CS	Czechoslovakia	LT	Lithuania	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	LV	Latvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FI	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia	UZ	Uzbekistan
GA	Gabon	MR	Mauritania	VN	Viet Nam

PRODUCTION AND USE OF TYPE 5 17BETA-HYDROXYSTEROID DEHYDROGENASE

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates to the isolation and characterization of a novel enzyme which is implicated in the production of sex steroids, and more particularly, to the characterization of the gene and cDNA of a novel 20∝, 17β-hydroxysteroid dehydrogenase (hereinafter type 5 17β-HSD) which has been implicated in the conversion of progesterone and 4-androstenedione (Δ^4-dione) to 20∝-hydroxyprogesterone (20∝-OH-P) and testosterone (T), respectively. The use of this enzyme as an assay for inhibitors of the enzyme is also described, as are several other uses of the DNA, fragments thereof and antisense fragments thereof.

Description of the Related Art

The enzymes identified as 17β-HSDs are important in the production of human sex steroids, including androst-5-ene-3β,17β-diol (Δ^5-diol), testosterone and estradiol. It was once thought that a single gene encoded a single type of 17β-HSD which was responsible for catalyzing all of the reactions. However, in humans, several types of 17β-HSD have now been identified and characterized. Each type of 17β-HSD has been found to catalyze specific reactions and to be located in specific tissues. Further information about Types 1, 2 and 3 17β-HSD can be had by reference as follows: Type 1 17β-HSD is described by Luu-The, V. et al., *Mol. Endocrinol.*, 3:1301-1309 (1989) and by Peltoketo, H. et al., *FEBS Lett*, 239:73-77 (1988); Type 2 17β-HSD is described in Wu, L. et al., *J. Biol Chem*, 268:12964-12969 (1993); Type 3 17β-HSD is described in Geissler, WM, *Nature Genetics*, 7:34-39 (1994). A fourth type which is homologous to porcine ovarian 17β-HSD has recently been identified by researchers Adamski and de Launoit, however, applicant is not presently aware of published information on this type.

The present invention relates to a fifth type of 17β-HSD which is described in detail below.

SUMMARY OF THE INVENTION

It is an object of the present invention to provide a novel 17β-hydroxysteroid dehydrogenase (17β-HSD) which is identified as type 5 17β-HSD.

It is also an object of the present invention to provide a 17β-HSD which has been shown to be involved in the conversion of Δ^4-dione to testosterone and in the conversion of progesterone to 20∝-hydroxyprogesterone (20∝-OH-P).

It is a further object of this invention to provide the nucleotide sequences and a gene map for type 5 17β-HSD.

It is also an object of this invention to provide methods of using type 5 17β-HSD in an assay to identify compounds which inhibit the activity of this enzyme, and thus may reduce production of testosterone or 20∝-hydroxyprogesterone, and can be used to treat medical conditions which respond unfavorably to these steroids.

It is an additional object of this invention to provide methods of preventing the synthesis of type 5 17β-HSD by administering an antisense RNA of the gene sequence of type 5 17β-HSD to interfere with the translation of the gene's mRNA.

These and other objects are discussed herein.

In particular, a novel enzyme, type 5 17β-hydroxysteroid dehydrogenase, has been identified and characterized. The gene sequence for this type 5 17β-HSD was found to encode a protein of 323 amino acids, having an apparent calculated molecular weight of 36,844 daltons. The protein is encoded by nucleotides +11 through 982, including the stop codon (amino acids +1 through 323), numbered in the 5' to 3' direction, in the following sequence (SEQ ID Nos. 1 and 2):

```
GTGACAGGGA ATG GAT TCC AAA CAG CAG TGT GTA AAG CTA AAT GAT GGC   49
      Met Asp Ser Lys Gln Gln Cys Val Lys Leu Asn Asp Gly
        1       5          10

CAC TTC ATG CCT GTA TTG GGA TTT GGC ACC TAT GCA CCT CCA GAG GTT   97
His Phe Met Pro Val Leu Gly Phe Gly Thr Tyr Ala Pro Pro Glu Val
  15          20          25

CCG AGA AGT AAA GCT TTG GAG GTC ACC AAA TTA GCA ATA GAA GCT GGG  145
Pro Arg Ser Lys Ala Leu Glu Val Thr Lys Leu Ala Ile Glu Ala Gly
30          35          40          45

TTC CGC CAT ATA GAT TCT GCT CAT TTA TAC AAT AAT GAG GAG CAG GTT  193
```

Phe Arg His Ile Asp Ser Ala His Leu Tyr Asn Asn Glu Glu Gln Val
50 55 60

GGA CTG GCC ATC CGA AGC AAG ATT GCA GAT GGC AGT GTG AAG AGA GAA 241
Gly Leu Ala Ile Arg Ser Lys Ile Ala Asp Gly Ser Val Lys Arg Glu
65 70 75

GAC ATA TTC TAC ACT TCA AAG CTT TGG TCC ACT TTT CAT CGA CCA GAG 289
Asp Ile Phe Tyr Thr Ser Lys Leu Trp Ser Thr Phe His Arg Pro Glu
80 85 90

TTG GTC CGA CCA GCC TTG GAA AAC TCA CTG AAA AAA GCT CAA TTG GAC 337
Leu Val Arg Pro Ala Leu Glu Asn Ser Leu Lys Lys Ala Gln Leu Asp
95 100 105

TAT GTT GAC CTC TAT CTT ATT CAT TCT CCA ATG TCT CTA AAG CCA GGT 385
Tyr Val Asp Leu Tyr Leu Ile His Ser Pro Met Ser Leu Lys Pro Gly
110 115 120 125

GAG GAA CTT TCA CCA ACA GAT GAA AAT GGA AAA GTA ATA TTT GAC ATA 433
Glu Glu Leu Ser Pro Thr Asp Glu Asn Gly Lys Val Ile Phe Asp Ile
130 135 140

GTG GAT CTC TGT ACC ACC TGG GAG GCC ATG GAG AAG TGT AAG GAT GCA 481
Val Asp Leu Cys Thr Thr Trp Glu Ala Met Glu Lys Cys Lys Asp Ala
145 150 155

GGA TTG GCC AAG TCC ATT GGG GTG TCA AAC TTC AAC CGC AGG CAG CTG 529
Gly Leu Ala Lys Ser Ile Gly Val Ser Asn Phe Asn Arg Arg Gln Leu
160 165 170

GAG ATG ATC CTC AAC AAG CCA GGA CTC AAG TAC AAG CCT GTC TGC AAC 577
Glu Met Ile Leu Asn Lys Pro Gly Leu Lys Tyr Lys Pro Val Cys Asn
175 180 185

CAG GTA GAA TGT CAT CCG TAT TTC AAC CGG AGT AAA TTG CTA GAT TTC 625
Gln Val Glu Cys His Pro Tyr Phe Asn Arg Ser Lys Leu Leu Asp Phe
190 195 200 205

TGC AAG TCG AAA GAT ATT GTT CTG GTT GCC TAT AGT GCT CTG GGA TCT 673
Cys Lys Ser Lys Asp Ile Val Leu Val Ala Tyr Ser Ala Leu Gly Ser
210 215 220

CAA CGA GAC AAA CGA TGG GTG GAC CCG AAC TCC CCG GTG CTC TTG GAG 721
Gln Arg Asp Lys Arg Trp Val Asp Pro Asn Ser Pro Val Leu Leu Glu
225 230 235

GAC CCA GTC CTT TGT GCC TTG GCA AAA AAG CAC AAG CGA ACC CCA GCC 769
Asp Pro Val Leu Cys Ala Leu Ala Lys Lys His Lys Arg Thr Pro Ala
240 245 250

CTG ATT GCC CTG CGC TAC CAG CTG CAG CGT GGG GTT GTG GTC CTG GCC 817
Leu Ile Ala Leu Arg Tyr Gln Leu Gln Arg Gly Val Val Val Leu Ala
255 260 265

AAG AGC TAC AAT GAG CAG CGC ATC AGA CAG AAC GTG CAG GTT TTT GAG 865
Lys Ser Tyr Asn Glu Gln Arg Ile Arg Gln Asn Val Gln Val Phe Glu
270 275 280 285

TTC CAG TTG ACT GCA GAG GAC ATG AAA GCC ATA GAT GGC CTA GAC AGA 913
Phe Gln Leu Thr Ala Glu Asp Met Lys Ala Ile Asp Gly Leu Asp Arg
290 295 300

AAT CTC CAC TAT TTT AAC AGT GAT AGT TTT GCT AGC CAC CCT AAT TAT 961
Asn Leu His Tyr Phe Asn Ser Asp Ser Phe Ala Ser His Pro Asn Tyr
305 310 315

CCA TAT TCA GAT GAA TAT TAA CATGGAGACT TTGCCTGATG ATGTCTACCA 1012
Pro Tyr Ser Asp Glu Tyr *
320

GAAGGCCCTG TGTGTGGATG GTGACGCAGA GGACGTCTCT ATGCCGGTGA CTGGACATAT 1072

CACCTCTACT TAAATCCGTC CTGTTTAGCG ACTTCAGTCA ACTACAGCTC ACTCCATAGG 1132

CCAGAAATAC AATAAATCCT GTTTAGCGAC TTCAGTCAAC TACAGCTCAC TCCATAGGCC 1192

AGAAATACAA TAAA 1206

In addition, a complete gene map (Figure 5) and nucleotide sequences (SEQ. ID Nos. 3 through 29 and Figures 6A and 6B) of the chromosomal DNA of type 5 17β-HSD are provided. A more detailed description of the sequences will be provided *infra*.

The present invention includes methods for the synthetic production of type 5 17β-HSD, as well as peptides that are biologically functionally equivalent, and to methods of using these compounds to screen test compounds for their ability to inhibit or alter the enzymatic function. In addition, methods of producing antisense constructs to the type 5 17β-HSD gene's DNA or mRNA or portions thereof, and the use of those antisense constructs to interfere with the transcription or translation of the enzyme are also provided.

The nucleotide sequence which encodes type 5 17β-HSD and recombinant expression vectors which include the sequence may be modified so long as they continue to encode a functionally equivalent enzyme. Moreover, it is contemplated, within the invention, that codons within the coding region may be altered, *inter alia*, in a manner which, given the degeneracy of the genetic code, continues to encode the same protein or one providing a functionally equivalent protein. It is believed that nucleotide sequences analogous to SEQ ID No. 1, or those that hybridize under stringent conditions to the coding region of SEQ ID No. 1, are likely to encode a type 5 17β-HSD functionally equivalent to that encoded by the coding region of SEQ ID No. 1, especially if such analogous nucleotide sequence is at least 700, preferably at

least 850 and most preferably at least 969 nucleotides in length. As used herein, except where otherwise specified, "stringent conditions" means 0.1x SSC (0.3 M sodium chloride and 0.03M sodium citrate) and 0.1% sodium dodecyl sulphate (SDS) and 60° C.

It is also likely that tissues or cells from human or non-human sources and which tissues or cells have the enzymatic machinery to convert Δ^4-dione to testosterone, or to convert progesterone to 20α-hydroxyprogesterone, include a type 5 17β-HSD sufficiently analogous to human type 5 17β-HSD to be used in accordance with the present invention. In particular, cDNA libraries prepared from cells performing the foregoing conversions may be screening with probes in accordance with well known techniques prepared by reference to the nucleotides disclosed herein, and under varying degrees of stringency, in order to identify analogous cDNAs in other species. These analogous cDNAs are preferably at least 70% homologous to SEQ ID No. 1, more preferably at least 80% homologous, and most preferably at least 90% homologous. They preferably include stretches of perfect identity at least 10 nucleotides long, more preferably stretches of 15, 20 or even 30 nucleotides of perfect identity. Appropriate probes may be prepared from SEQ ID No. 1 or fragments thereof of suitable length, preferably at least 15 nucleotides in length. Confirmation with at least two distinct probes is preferred. Alternative isolation strategies, such as polymerase chain reaction (PCR) amplification, may also be used.

Homologous type 5 17β-HSDs so obtained, as well as the genes encoding them, are used in accordance with the invention in all of the ways for using SEQ ID No. 2 and SEQ ID No. 1, respectively.

Recombinant expression vectors can include the entire coding region for human type 5 17β-HSD as shown in SEQ ID No. 1, the coding region for human type 5 17β-HSD which has been modified, portions of the coding region for human type 5 17β-HSD, the chromosomal DNA of type 5 17β-HSD, an antisense construct to type 5 17β-HSD, or portions of antisense constructs to type 5 17β-HSD.

In the context of the invention, "isolated" means having a higher purity than exists in nature, but does not require purification from a natural source. Isolated nucleotides encoding type 5 17β-HSD may be produced synthetically, or by isolating cDNA thereof from a cDNA library or by any of numerous other methods well understood in the art.

In one embodiment, the invention provides an isolated nucleotide sequence encoding type 5 17β-hydroxysteroid dehydrogenase, said sequence being sufficiently homologous to SEQ ID No. 1 or a complement thereof, to hybridize under stringent conditions to the coding region of SEQ ID No. 1 or a complement thereof and said sequence encoding an enzyme which catalyzes the conversion of progesterone to 20∝-hydroxyprogesterone and the conversion of 4-androstenedione to testosterone.

In a further embodiment, the invention provides an isolated nucleotide sequence comprising at least ten consecutive nucleotides identical to 10 consecutive nucleotides in the coding region of SEQ ID No. 1, or the complement thereof.

In an additional embodiment, the invention provides an oligonucleotide sequence selected from the group consisting of SEQ ID Nos. 30 through 59.

In another embodiment, the invention provides a recombinant expression vector comprising a promoter sequence and an oligonucleotide sequence selected from the group of SEQ ID Nos. 30 to 59.

In a further embodiment, the invention provides a method of blocking synthesis of type 5 17β-HSD, comprising the step of introducing an oligonucleotide selected from the group consisting of SEQ ID Nos. 30 to 59 into cells.

In an additional embodiment, the invention provides an isolated chromosomal DNA fragment which upon transcription and translation encodes type 5 17β-hydroxysteroid dehydrogenase and wherein said fragment contains nine exons and wherein said fragment includes introns which are 16 kilobase pairs in length.

In another embodiment, the invention provides an isolated DNA sequence encoding type 5 17β-hydroxysteroid dehydrogenase, said sequence being sufficiently homologous to SEQ ID No. 3 or a complement thereof, to hybridize under stringent conditions to SEQ ID No. 3, or its complement.

In a further embodiment, the invention provides a method for producing type 5 17β-hydroxysteroid dehydrogenase, comprising the steps of preparing a recombinant host transformed or transfected with the vector of claim 3 and culturing said host under conditions which are conducive to the production of type 5 17β-hydroxysteroid dehydrogenase by said host.

In an additional embodiment, the invention provides a method for determining the inhibitory effect of a test compound on the enzymatic activity of type 5 17β-hydroxysteroid dehydrogenase, comprising the steps of providing type 5 17β-

hydroxysteroid dehydrogenase; contacting said type 5 17β-hydroxysteroid dehydrogenase with said test compound; and thereafter determining the enzymatic activity of said type 5 17β-hydroxysteroid dehydrogenase in the presence of said test compound.

In an additional embodiment, the invention provides a method of interfering with the expression of type 5 17β-hydroxysteroid dehydrogenase, comprising the step of administering nucleic acids substantially identical to at least 15 consecutive nucleotides of SEQ ID No. 1 or a complement thereof.

In a further embodiment, there is provided a method of interfering with the synthesis of type 5 17β-hydroxysteroid dehydrogenase, comprising the step of administering antisense RNA complementary to mRNA encoded by at least 15 consecutive nucleotides of SEQ ID No. 1 or a complement thereof.

In an additional embodiment, the invention provides a method of interfering with the expression of type 5 17β-hydroxysteroid dehydrogenase, comprising the step of administering nucleic acids substantially identical to at least 15 consecutive nucleotides of SEQ ID No. 3 or a complement thereof.

In another embodiment, the invention provides a method of interfering with the synthesis of type 5 17β-hydroxysteroid dehydrogenase, comprising the step of administering antisense RNA complementary to mRNA encoded by at least 15 consecutive nucleotides of SEQ ID No. 3 or a complement thereof.

In a further embodiment, there is provided a method for determining the inhibitory effect of antisense nucleic acids on the enzymatic activity of type 5 17β-hydroxysteroid dehydrogenase, comprising the steps of providing a host system capable of expressing type 5 17β-hydroxysteroid dehydrogenase; introducing said antisense nucleic acids into said host system; and thereafter determining the enzymatic activity of said type 5 17β-hydroxysteroid dehydrogenase.

Other features and advantages of the present invention will become apparent from the following description of the invention which refers to the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A and 1B are graphs showing the enzymatic activities of Type 5 17β-

HSD on various substrates. The enzyme was expressed in embryonal kidney (293) cells (ATCC CRL 1573) which were transfected with a vector, prepared in accordance with the invention, and containing the gene encoding human type 5 17β-HSD. Figure 1A shows the substrate specificity of type 5 17β-HSD. The concentration of each substrate was 0.1 μM. Figure 1B shows the time course amount of 20∝-HSD and 17β-HSD activities of cells transfected with vectors containing human type 5 17β-HSD. The substrates, progesterone (P) and Δ^4-dione, were added at a concentration of 0.1 μM;

Figure 2 is a map of a pCMV vector which is exemplary of one that can be used to transfect host cells in accordance with the invention;

Figure 3 is the cDNA sequence (SEQ ID No. 1) and the deduced amino acid sequence (SEQ ID No. 2) of human type 5 17β-HSD. The nucleotide sequence is numbered in the 5' to 3' direction with the adenosine of the initiation codon (ATG) designated as +11. The translation stop codon is indicated by asterisks. The potential post modification sites are underlined, wherein TSK = tyrosine sulfokinase; CK2 = casein kinase II; PKC = protein kinase C; NG = N-glycosylation; and NM = N-myrystoylation;

Figure 4 is a comparison of the deduced amino acid sequence of human type 5 17β-HSD to the amino acid sequences of rabbit (rb), rat (r), and bovine (b) 20∝-HSD as well as human (h) and rat (r) 3∝-HSD, bovine prostaglandin f synthase (b pgfs) and frog ρ-crystallin (f ρ-crys). The amino sequences are indicated using the conventional single letter code and are numbered on the right. The dashes (-) and dots (.) indicate identical and missing amino acid residues, respectively;

Figure 5 is a map of the chromosomal DNA of a gene which encodes type 5 17β-HSD; and

Figures 6A and 6B are the nucleotide sequence of the chromosomal DNA of a gene which encodes type 5 17β-HSD.

DETAILED DESCRIPTION OF THE INVENTION

A gene encoding the enzyme, type 5 17β-HSD, has been isolated and encodes a protein having 323 amino acids with a calculated molecular weight of 36,844 daltons. As shown in Figure 3, the coding portion of this gene includes nucleotides +11 through 982, including the stop codon (and encodes amino acids +1 through 323), numbered in the 5' to 3' direction.

The chromosomal DNA fragment of the gene for type 5 17β-HSD has also been characterized. A map of the gene is provided in Figure 5. In particular, it was found, using primer extension analysis, that the gene includes 16 kilobase pairs (kb) and contained nine short exons. A portion of the 5' flanking region, as set forth in SEQ ID No. 3, of the genomic DNA includes 730 base pairs (bp). Exon I (SEQ ID No. 4) contains 37 nucleotides in the 5'-noncoding region and the nucleotides for the first 28 amino acids. The second intron region includes the nucleotides set forth in SEQ ID Nos. 5 and 6, which are 252 and 410 bp, respectively. These are joined by a 1.2 kb region which is not important and therefore, its sequence has been omitted. Exon 2 (SEQ ID No. 7) contains nucleotides for the following 56 amino acids of human type 5 17β-HSD. The following intron region includes SEQ ID Nos. 8 and 9, 700 and 73 bp, respectively, which are joined by a 0.1 kb region for which the sequence has not been provided. Exon 3 (SEQ ID No. 10) includes the next 117 nucleotides which specify the following 39 amino acids. The fourth intron region is represented by SEQ ID Nos. 11 and 12, 152 and 208 nucleotides in length, respectively, with a 0.9 kb region in between which has not been provided. Exon 4 (SEQ ID No. 13) includes the next 78 bp which specify the following 26 amino acids of the enzyme. Intron region five contains SEQ ID Nos. 14 and 15, with 98 and 249 nucleotides, respectively, with a 0.1 kb region in the middle which has not been provided. The fifth exon (SEQ ID No. 16) contains nucleotides for the following 41 amino acids of human type 5 17β-HSD. The sixth intron region, set forth in SEQ ID Nos. 17 and 18 with 138 and 189 bp, respectively, also includes a 2.8 kb region which has not been provided. Exon 6 (SEQ ID No. 19) contains nucleotides for the following 36 amino acids of type 5 17β-HSD, as well as two nucleotides of the codon 227 (Trp). The next intron region includes a 136 bp portion (SEQ ID No. 20) and a

66 bp portion (SEQ ID No. 21) which are joined by a 0.1 kb region which is not set forth. Exon 7 (SEQ ID No. 22) contains nucleotides for the third nucleotide of codon 227 (Trp) and nucleotides for the following 55 codons. The following intron region includes a 136 nucleotide region (SEQ ID No. 23), a 2.5 kb region which is not provided and a 286 bp region (SEQ ID No. 24). Exon 8 (SEQ ID No. 25) includes 83 nucleotides which code for the following 27 amino acids and 2 nucleotides of codon 310. The ninth intron region contains 713 nucleotides (SEQ ID No. 26) followed by a 1 kb region which has not been provided followed by a 415 nucleotide region (SEQ ID No. 27). Exon 9 (SEQ ID No. 28) contains the third nucleotide of codon 310, 42 nucleotides for the last 13 amino acids and a stop codon and approximately 200 nucleotides in the 3'-untranslated region. A polymorphic $(GT)_n$ repeat region that can be used to perform genetic linkage mapping of the type 5 17β-HSD can be found 255 nucleotides downstream from the TAA stop codon. SEQ ID No. 29 sets forth 109 bp of additional genomic sequence. The nucleotide sequence of the gene fragment, as described above, is provided in Figures 6A and 6B.

The type 5 17β-HSD enzyme can be produced by incorporating the nucleotide sequence for the coding portion of the gene into a vector which is then transformed or transfected into a host system which is capable of expressing the enzyme. The DNA can be maintained transiently in the host or can be stably integrated into the genome of the host cell. In addition, the chromosomal DNA can be incorporated into a vector and transfected into a host system for cloning.

In particular, for the cloning and expression of type 5 17β-HSD, any common expression vectors, such as plasmids, can be used. These vectors can be prokaryotic expression vectors including those derived from bacteriophage λ such as λgt11 and λEMBL3, *E. coli* strains such as pBR322 and Bluescript (Stratagene); or eukaryotic vectors, such as those in the pCMV family. Vectors incorporating an isolated human cDNA shown in Sequence ID No. 1 (ATCC Deposit No.) and the chromosomal DNA as shown in Sequence ID Nos. 3 through 29 (ATCC Deposit No.) for type 5 17β-HSD have been placed on deposit at the American Type Culture Collection (ATCC, Rockville, MD), in accordance with the terms of the Budapest Treaty, and will be made available to the public upon issuance of a patent based on the present patent application.

These vectors generally include appropriate replication and control sequences

which are compatible with the host system into which the vectors are transfected. A promoter sequence is generally included. For prokaryotes, some representative promoters include β-lactamase, lactose, and tryptophan. In mammalian cells, commonly used promoters include, but are not limited to, adenovirus, cytomegalovirus (CMV) and simian virus 40 (SV40). The vector can also optionally include, as appropriate, an origin of replication, ribosome binding sites, RNA splice sites, polyadenylation sites, transcriptional termination sequences and/or a selectable marker. It is well understood that there are a variety of vector systems with various characteristics which can be used in the practice of the invention. A map of the pCMV vector, which is an example of a vector which can be used in the practice of the invention, is provided in Figure 2.

Commonly known host systems which are known for expressing an enzyme, and which may be transfected with an appropriate vector which includes a gene for Type 5 17β-HSD can be used in the practice of the invention. These host systems include prokaryotic hosts, such as *E. coli*, bacilli, such as *Bacillus subtilus*, and other enterobacteria, such as *Salmonella*, *Serratia*, and *Pseudomonas* species. Eukaryotic microbes, including yeast cultures, can also be used. The most common of these is *Saccharomyces cerevisiae*, although other species are commercially available and can be used. Furthermore, cell cultures can be grown which are derived from mammalian cells. Some examples of suitable host cell lines include embryonal kidney (293), SW-13, chinese hamster ovary (CHO), HeLa, myeloma, Jurkat, COS-1, BHK, W138 and madin-darby canine kidney (MDCK). In the practice of the invention, the 293 cells are preferred.

Type 5 17β-HSD, whether recombinantly produced as described herein, purified from nature, or otherwise produced, can be used in assays to identify compounds which inhibit or alter the activity of the enzyme. In particular, since type 5 17β-HSD is shown to catalyze the conversion of progesterone to 20α-OH-P and the conversion of Δ^4-dione to testosterone, this enzyme can be used to identify compounds which interfere with the production of these sex steroids. It is preferred that the enzyme be obtained directly from the recombinant host, wherein following expression, a crude homogenate is prepared which includes the enzyme. A substrate of the enzyme, such as progesterone or Δ^4-dione and a compound to be tested are then mixed with the homogenate. The activity of the enzyme with and without the test compound

is compared. Numerous methods are known which can be used to indicate the effects of the test compound on the activity of the substrate for easy detection of the relative amounts of substrate and product over time. For example, it is possible to label the substrate so that the label also stays on any product that is formed. Radioactive labels, such as C^{14} or H^3, which can be quantitatively analyzed are particularly useful.

It is preferred that the mixture of the enzyme, test compound and substrate be allowed to incubate for a predetermined amount of time. In addition, it is preferred that the product is separated from the substrate for easier analysis. A number of separation techniques are known, for example, thin layer chromatography (TLC), high pressure liquid chromatography (HPLC), spectrophotometry, gas chromatography, mass spectrophotometry and nuclear magnetic resonance (NMR). However, any known method which can differentiate between a substrate and a product can be used.

It is also contemplated that the gene for type 5 17β-HSD or a portion thereof can be used to produce antisense nucleic acid sequences for inhibiting expression of Type 5 17β-HSD *in vivo*. Thus activity of the enzyme and levels of its products (e.g. testosterone) may be reduced where desirable. In general, antisense nucleic acid sequences can interfere with transcription, splicing or translation processes. Antisense sequences can prevent transcription by forming a triple helix or hybridizing to an opened loop which is created by RNA polymerase or hybridizing to nascent RNA. On the other hand, splicing can advantageously be interfered with if the antisense sequences bind at the intersection of an exon and an intron. Finally, translation can be affected by blocking the binding of initiation factors or by preventing the assembly of ribosomal subunits at the start codon or by blocking the ribosome from the coding portion of the mRNA, preferably by using RNA that is antisense to the message. For further general information, see Hélène et al., *Biochimica et Biophysica Acta*, 1049:99-125 (1990), which is herein incorporated by reference in its entirety.

An antisense nucleic acid sequence is an RNA or single stranded DNA sequence which is complementary to the target portion of the target gene. These antisense sequences are introduced into cells where the complementary strand base pairs with the target portion of the target gene, thereby blocking the transcription, splicing or translation of the gene and eliminating or reducing the production of type 5 17β-HSD. The length of the antisense nucleic acid sequence need be no more than is sufficient to interfere with the transcription, splicing or translation of functional type 5

17β-HSD. Antisense strands can range in size from 10 nucleotides to the complete gene, however, about 10 to 50 nucleotides are preferred, and 15 to 25 nucleotides are most preferred.

Although it is contemplated that any portion of the gene could be used to produce antisense sequences, it is preferred that the antisense is directed to the coding portion of the gene or to the sequence around the translation initiation site of the mRNA or to a portion of the promoter. Some examples of specific antisense oligonucleotide sequences in the coding region which can be used to block type 5 17β-HSD synthesis include: TTTAGCTTTACACACTGCTGTT (SEQ ID No. 30); TCCAAAGCTTTACTTCTCGG (SEQ ID No. 31); GATGAAAAGTGGACCA (SEQ ID No. 32); ATCTGTTGGTGAAAGTTC (SEQ ID No. 33); TCCAGCTGCCTGCGGT (SEQ ID No. 34); CTTGTACTTGAGTCCTG (SEQ ID No. 35); CTCCGGTTGAAATACGGA (SEQ ID No. 36); CATCGTTTGTCTCGTTGAGA (SEQ ID No. 37); TCACTGTTAAAATAGTGGAGAT (SEQ ID No. 38); ATCTGAATATGGATAAT (SEQ ID No. 39). Examples of antisense oligonucleotide sequences which can block the splicing of the type 5 17β-HSD premessage are as follows: TTCTCGGAACCTGGAGGAGC (SEQ ID No. 40); GACACAGTACCTTTGAAGTG (SEQ ID No. 41); TGGACCAAAGCTGCAGAGGT (SEQ ID No. 42); CCTCACCTGGCTGAAATAGA (SEQ ID No. 43); AAGCACTCACCTCCCAGGTG (SEQ ID No. 44); GACATTCTACCTGCAGTTGA (SEQ ID No. 45); CTCAAAAACCTATCAGAAA (SEQ ID No. 46); GGAAACTTACCTATCACTGT (SEQ ID No. 47); GCTAGCAAAACTGAAAAGAG (SEQ ID No. 48). Examples of antisense oligonucleotide sequences which inhibit the promoter activity of type 5 17β-HSd include: GAGAAATATTCATTCTG (SEQ ID No. 49); CGAGTCCTGATAAAGCTG (SEQ ID No. 50); GATGAGGGTGCAAATAA (SEQ ID No. 51); GGAGTGTTAATTAATAACAGTTT (SEQ ID No. 52); CAGAGATTACAAAAACAAT (SEQ ID No. 53); TGCCTTTTTACATTTTCAATCA (SEQ ID No. 54); ACACATAATTTAAAGGA (SEQ ID No. 55); TTAAATTATTCAAAAGG (SEQ ID No. 56); AAGAGAAATATTCATTTCTG (SEQ ID No. 57);

CCCCTCCCCCCACCCCTGCA (SEQ ID No. 58); CTGCCGTGATAATGCCCC (SEQ ID No. 59).

As is well understood in the art, the oligonucleotide sequences can be modified in various manners in order to increase the effectiveness of the treatment with oligonucleotides. In particular, the sequences can be modified to include additional RNA to the 3' end of the RNA which can form a hairpin-loop structure and thereby prevent degradation by nucleases. In addition, the chemical linkages in the backbone of the oligonucleotides can be modified to also prevent cleavage by nucleases.

There are numerous methods which are known in the art for introducing the antisense strands into cells. One strategy is to incorporate the gene which encodes type 5 17β-HSD in the opposite orientation in a vector so that the RNA which is transcribed from the plasmid is complementary to the mRNA transcribed from the cellular gene. A strong promoter, such as pCMV, is generally included in the vector, upstream of the gene sequence, so that a large amount of the antisense RNA is produced and is available for binding sense mRNA. The vectors are then transfected into cells which are then administered. It is also possible to produce single stranded DNA oligonucleotides or antisense RNA and incorporate these into cells or liposomes which are then administered. The use of liposomes, such as those described in WO95/03788, which is herein incorporated by reference, is preferred. However, other methods which are well understood in the art can also be used to introduce the antisense strands into cells and to administer to these patients in need of such treatment.

The following is an example of the expression of human type 5 17β-HSD. This example is intended to be illustrative of the invention and it is well understood by those of skill in the art that modifications, alterations and different techniques can be used within the scope of the invention.

Expression of

20∝, 17β-HSD (Type 5 17β-HSD)

Construction of the expression vector and nucleotide sequence determination

The phage DNA were digested with EcoRI restriction enzyme and the resulting cDNA fragments were inserted in the EcoRI site downstream to the cytomegalovirus

(CMV) promoter of the pCMV vector as shown in Figure 2. The recombinant pCMV plasmids were amplified in *Escherichia coli* DH5α competent cells, and were isolated using the alkaline lysis procedure as described by Maniatis in Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Press 1982). The sequencing of double-stranded plasmid DNA was performed according to the dideoxy chain termination method described by Sanger F. et al., *Proc. Natl. Acad. Sci.*, 74:5463-5467 (1977) using a T7 DNA polymerase sequencing kit (Pharmacia LKB Biotechnology). In order to avoid errors, all sequences were determined by sequencing both strands of the DNA. The oligonucleotide primers were synthesized using a 394 DNA/RNA synthesizer (Applied Biosystem).

As shown in Figure 2, the pCMV vector contains 582 nucleotides of the pCMV promoter, followed by 74 nucleotides of unknown origin which includes the EcoRI and HindIII sites, followed by 432 basepairs (bp) of a small t intron (fragment 4713 - 4570) and a polyadenylation signal (fragment 2825 - 2536) of SV40, followed by 156 nucleotides of unknown origin, followed by 1989 bp of the PvuII (628) to AatII (2617) fragment from the pUC 19 vector (New England Biolabs) which contains an *E. coli* origin of replication and an ampicillin resistance gene for propagation in *E. coli*.

Transient expression in transformed embryonal kidney (293) cells

The vectors were transfected using the calcium phosphate procedure described by Kingston, R.E., In: Current Protocols in Molecular Biology, Ausubel et al. eds., pp. 9.1.1 - 9.1.9, John Wiley & Sons, N.Y. (1987) and used 1 to 10 μg of recombinant plasmid DNA per 10^6 cells. The total amount of DNA is kept at 10μg of plasmid DNA per 10^6 cells by completing with pCMV plasmid without insert. The cells were initially plated at 10^4 cells/cm^2 in Falcon® culture flasks and grown in Dulbecco's modified Eagle's medium containing 10% (vol/vol) fetal bovine serum (hyclone, Logan, UT) under a humidified atmosphere of air/CO^2 (95%/5%) at 37°C and supplemented with 2 mM L-glutamine, 1 mM sodium pyruvate, 100 IU penicillin/ml, and 100 μg streptomycin sulfate/ml.

Assay of enzymatic activity

The determination of enzymatic activity was performed as described by Luu-

The et al., *Biochemistry*, 13:8861-8865 (1991) which is herein incorporated by reference. See also Lachance et al., *J. Biol. Chem.*, 265:20469 - 20475 (1990). Briefly, 0.1 µM of the indicated ^{14}C-labeled substrate (Dupont Inc. (Canada)), namely, dehydroepiandrosterone (DHEA), 4-androstene-3,17-dione (Δ^4-dione), testosterone (T), estrone (E1), estradiol (E2), dihydrotestosterone (DHT), and progesterone (PROG), was added to freshly changed culture medium in a 6-well culture plate. After incubation for 1 hour, the steroids were extracted twice with 2 ml of ether. The organic phase was pooled and evaporated to dryness. The steroids were solubilized in 50 µl of dichloromethane, applied to a Silica gel 60 thin layer chromatography (TLC) plate (Merck, Darmstad, Germany) and then separated by migration in the toluene-acetone (4:1) solvent system (Luu-The, V. et al., *J. Invest. Dermatol.*, 102:221-226 (1994) which is herein incorporated by reference). The substrates and metabolites were identified by comparison with reference steroids, revealed by autoradiography and quantitated using the Phosphoimager System (Molecular Dynamics, Sunnyvale, CA).

Cloning of the type 5 17β-HSD genomic DNA clone

The hybridization and sequencing methods were as described above and as previously described (Luu-The et al., *Mol. Endocrinol.*, 4:268-275 (1990); Luu-The et al., *DNA and Cell Biol.*, 14:511-518 (1995); Lachance et al., *J. Biol. Chem.*, 265:20469-20475 (1990); Lachance et al., *DNA and Cell Biol.* 10:701-711 (1991); Bernier et al., *J. Biol. Chem.* 269, 28200-28205, (1994) which are herein incorporated by reference).

About 20 recombinant clones which gave the strongest hybridization signal were selected for second and third screening in order to isolate a single phage plaque. The two longest clones that hybridized with specific oligonucleotides probes located at the 5' and 3' regions of type 5 17β-HSD, respectively, were selected for mapping, subcloning and sequencing. As shown in Figures 5 and 6, the gene is included in approximately 16 kilobase pairs of introns and contains 9 short exons. A primer extension analysis using oligoprimer CAT-CAT-TTA-GCT-TTA-CAT-ACT-GCT-G located at positions 13 to 27, indicates that the start site is situated 37 nucleotides upstream from the ATG initiating codon.

The sites and signatures in the primary protein sequence were detected using PC/Gene software (Intelli Genetics Inc., Mountain View, CA). This analysis revealed a potential N-glycosylation site at Asn-198; five protein kinase C sites at Ser-73, Thr-82, Ser-102, Ser-121, and Ser-221; five casein kinase II phosphorylation sites at Ser-129, Thr-146, Ser-221, Ser-271, and Thr-289; two N-myristoylation sites at Gly-158 and Gly-298; a tyrosine sulfatation site at Tyr-55; an aldo/keto reductase family signature 1 (25) at amino acids 158 to 168 and an aldo/keto reductase family putative active site signature at amino acids 262 to 280.

As described above, the enzymatic activity of the type 5 17β-HSD was evaluated by transfecting 293 cells with vectors which included the gene encoding human type 5 17β-HSD. The ability of the enzyme to catalyze the transformation of progesterone (P) to 20∝-hydroxyprogesterone (20∝-OH-P), 4-androstenedione (Δ^4-dione) to testosterone (T), 5∝-androstane-3,17-dione (A-dione) to dihydrotestosterone (DHT), dehydroepiandrosterone (DHEA) to 5-androstene-3β,17β-diol, and estrone (E1) to estradiol (E2) was analyzed. As shown in Figure 1A, the enzyme possesses high reductive 20∝-HSD activity, wherein progesterone (P) is transformed to the inactive 20∝-OH-P, and 17β-HSD activity, wherein Δ^4-dione is converted to testosterone (T). However, 3∝-HSD activity which is responsible for the transformation of DHT to 5α-androstane-3α,17β-diol is negligible. The ability of this enzyme to transform E1 and E2 was also negligible (Figure 1A). Figure 1B shows that the 20∝-HSD and 17β-HSD activities increased over time.

The isolated amino acid sequence of human type 5 17β-HSD was also compared with rabbit 20∝-HSD (rb), rat 20∝-HSD (r), human 3∝-HSD (h), rat 3∝-HSD (r), bovine prostaglandin f synthase (b pgfs), frog ρ-crystallin (f ρ-crys) and human type 1 and type 2 17β-HSDs (h) as shown in Figure 4. These sequences show 76.2%, 70.7%, 84.0%, 68.7%, 78.3%, 59.7%, 15.2% and 15.0% identity with type 5 17β-HSD, respectively.

Although the present invention has been described in relation to particular embodiments thereof, many other variations and modifications and other uses will be apparent to those skilled in the art.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: LUU-THE, Van
LABRIE, Fernand

(ii) TITLE OF INVENTION: PRODUCTION AND USE OF ISOLATED TYPE 5 17B-HYDROXYSTEROID DEHYDROGENASE

(iii) NUMBER OF SEQUENCES: 59

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
(B) STREET: 1180 Avenue of the Americas
(C) CITY: New York
(D) STATE: NY
(E) COUNTRY: US
(F) ZIP: 10036-8403

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Meilman, Edward
(B) REGISTRATION NUMBER: 24,735
(C) REFERENCE/DOCKET NUMBER: P/1259-313

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (212) 382-0700
(B) TELEFAX: (212) 382-0888
(C) TELEX: 236925

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1206 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 11..982

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
GTGACAGGGA ATG GAT TCC AAA CAG CAG TGT GTA AAG CTA AAT GAT GGC          49
           Met Asp Ser Lys Gln Gln Cys Val Lys Leu Asn Asp Gly
             1               5                  10

CAC TTC ATG CCT GTA TTG GGA TTT GGC ACC TAT GCA CCT CCA GAG GTT          97
His Phe Met Pro Val Leu Gly Phe Gly Thr Tyr Ala Pro Pro Glu Val
     15                  20                  25

CCG AGA AGT AAA GCT TTG GAG GTC ACC AAA TTA GCA ATA GAA GCT GGG         145
Pro Arg Ser Lys Ala Leu Glu Val Thr Lys Leu Ala Ile Glu Ala Gly
 30                  35                  40                  45

TTC CGC CAT ATA GAT TCT GCT CAT TTA TAC AAT AAT GAG GAG CAG GTT         193
Phe Arg His Ile Asp Ser Ala His Leu Tyr Asn Asn Glu Glu Gln Val
                 50                  55                  60

GGA CTG GCC ATC CGA AGC AAG ATT GCA GAT GGC AGT GTG AAG AGA GAA         241
Gly Leu Ala Ile Arg Ser Lys Ile Ala Asp Gly Ser Val Lys Arg Glu
             65                  70                  75

GAC ATA TTC TAC ACT TCA AAG CTT TGG TCC ACT TTT CAT CGA CCA GAG         289
Asp Ile Phe Tyr Thr Ser Lys Leu Trp Ser Thr Phe His Arg Pro Glu
         80                  85                  90

TTG GTC CGA CCA GCC TTG GAA AAC TCA CTG AAA AAA GCT CAA TTG GAC         337
Leu Val Arg Pro Ala Leu Glu Asn Ser Leu Lys Lys Ala Gln Leu Asp
     95                  100                 105

TAT GTT GAC CTC TAT CTT ATT CAT TCT CCA ATG TCT CTA AAG CCA GGT         385
Tyr Val Asp Leu Tyr Leu Ile His Ser Pro Met Ser Leu Lys Pro Gly
110                 115                 120                 125

GAG GAA CTT TCA CCA ACA GAT GAA AAT GGA AAA GTA ATA TTT GAC ATA         433
Glu Glu Leu Ser Pro Thr Asp Glu Asn Gly Lys Val Ile Phe Asp Ile
                130                 135                 140

GTG GAT CTC TGT ACC ACC TGG GAG GCC ATG GAG AAG TGT AAG GAT GCA         481
Val Asp Leu Cys Thr Thr Trp Glu Ala Met Glu Lys Cys Lys Asp Ala
            145                 150                 155

GGA TTG GCC AAG TCC ATT GGG GTG TCA AAC TTC AAC CGC AGG CAG CTG         529
Gly Leu Ala Lys Ser Ile Gly Val Ser Asn Phe Asn Arg Arg Gln Leu
        160                 165                 170

GAG ATG ATC CTC AAC AAG CCA GGA CTC AAG TAC AAG CCT GTC TGC AAC         577
Glu Met Ile Leu Asn Lys Pro Gly Leu Lys Tyr Lys Pro Val Cys Asn
    175                 180                 185

CAG GTA GAA TGT CAT CCG TAT TTC AAC CGG AGT AAA TTG CTA GAT TTC         625
Gln Val Glu Cys His Pro Tyr Phe Asn Arg Ser Lys Leu Leu Asp Phe
190                 195                 200                 205

TGC AAG TCG AAA GAT ATT GTT CTG GTT GCC TAT AGT GCT CTG GGA TCT         673
Cys Lys Ser Lys Asp Ile Val Leu Val Ala Tyr Ser Ala Leu Gly Ser
                210                 215                 220

CAA CGA GAC AAA CGA TGG GTG GAC CCG AAC TCC CCG GTG CTC TTG GAG         721
Gln Arg Asp Lys Arg Trp Val Asp Pro Asn Ser Pro Val Leu Leu Glu
            225                 230                 235

GAC CCA GTC CTT TGT GCC TTG GCA AAA AAG CAC AAG CGA ACC CCA GCC         769
Asp Pro Val Leu Cys Ala Leu Ala Lys Lys His Lys Arg Thr Pro Ala
        240                 245                 250
```

```
CTG ATT GCC CTG CGC TAC CAG CTG CAG CGT GGG GTT GTG GTC CTG GCC          817
Leu Ile Ala Leu Arg Tyr Gln Leu Gln Arg Gly Val Val Val Leu Ala
    255                 260                 265

AAG AGC TAC AAT GAG CAG CGC ATC AGA CAG AAC GTG CAG GTT TTT GAG          865
Lys Ser Tyr Asn Glu Gln Arg Ile Arg Gln Asn Val Gln Val Phe Glu
270                 275                 280                 285

TTC CAG TTG ACT GCA GAG GAC ATG AAA GCC ATA GAT GGC CTA GAC AGA          913
Phe Gln Leu Thr Ala Glu Asp Met Lys Ala Ile Asp Gly Leu Asp Arg
                290                 295                 300

AAT CTC CAC TAT TTT AAC AGT GAT AGT TTT GCT AGC CAC CCT AAT TAT          961
Asn Leu His Tyr Phe Asn Ser Asp Ser Phe Ala Ser His Pro Asn Tyr
            305                 310                 315

CCA TAT TCA GAT GAA TAT TAA CATGGAGACT TTGCCTGATG ATGTCTACCA             1012
Pro Tyr Ser Asp Glu Tyr *
        320

GAAGGCCCTG TGTGTGGATG GTGACGCAGA GGACGTCTCT ATGCCGGTGA CTGGACATAT       1072

CACCTCTACT TAAATCCGTC CTGTTTAGCG ACTTCAGTCA ACTACAGCTC ACTCCATAGG       1132

CCAGAAATAC AATAAATCCT GTTTAGCGAC TTCAGTCAAC TACAGCTCAC TCCATAGGCC       1192

AGAAATACAA TAAA                                                        1206
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 324 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Asp Ser Lys Gln Gln Cys Val Lys Leu Asn Asp Gly His Phe Met
 1               5                  10                  15

Pro Val Leu Gly Phe Gly Thr Tyr Ala Pro Pro Glu Val Pro Arg Ser
            20                  25                  30

Lys Ala Leu Glu Val Thr Lys Leu Ala Ile Glu Ala Gly Phe Arg His
        35                  40                  45

Ile Asp Ser Ala His Leu Tyr Asn Asn Glu Glu Gln Val Gly Leu Ala
    50                  55                  60

Ile Arg Ser Lys Ile Ala Asp Gly Ser Val Lys Arg Glu Asp Ile Phe
65                  70                  75                  80

Tyr Thr Ser Lys Leu Trp Ser Thr Phe His Arg Pro Glu Leu Val Arg
                85                  90                  95

Pro Ala Leu Glu Asn Ser Leu Lys Lys Ala Gln Leu Asp Tyr Val Asp
            100                 105                 110

Leu Tyr Leu Ile His Ser Pro Met Ser Leu Lys Pro Gly Glu Glu Leu
        115                 120                 125

Ser Pro Thr Asp Glu Asn Gly Lys Val Ile Phe Asp Ile Val Asp Leu
    130                 135                 140
```

Cys Thr Thr Trp Glu Ala Met Glu Lys Cys Lys Asp Ala Gly Leu Ala
145 150 155 160

Lys Ser Ile Gly Val Ser Asn Phe Asn Arg Arg Gln Leu Glu Met Ile
165 170 175

Leu Asn Lys Pro Gly Leu Lys Tyr Lys Pro Val Cys Asn Gln Val Glu
180 185 190

Cys His Pro Tyr Phe Asn Arg Ser Lys Leu Leu Asp Phe Cys Lys Ser
195 200 205

Lys Asp Ile Val Leu Val Ala Tyr Ser Ala Leu Gly Ser Gln Arg Asp
210 215 220

Lys Arg Trp Val Asp Pro Asn Ser Pro Val Leu Leu Glu Asp Pro Val
225 230 235 240

Leu Cys Ala Leu Ala Lys Lys His Lys Arg Thr Pro Ala Leu Ile Ala
245 250 255

Leu Arg Tyr Gln Leu Gln Arg Gly Val Val Val Leu Ala Lys Ser Tyr
260 265 270

Asn Glu Gln Arg Ile Arg Gln Asn Val Gln Val Phe Glu Phe Gln Leu
275 280 285

Thr Ala Glu Asp Met Lys Ala Ile Asp Gly Leu Asp Arg Asn Leu His
290 295 300

Tyr Phe Asn Ser Asp Ser Phe Ala Ser His Pro Asn Tyr Pro Tyr Ser
305 310 315 320

Asp Glu Tyr *

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 730 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

AAGAACAAAT ACTATTAAGG CACTGCTTGC ATATATTAAA TGATGTCCAA ACTCCAAAAA 60

CTGTTAATAA TTAACACTCC AATAAAAACT ACACCAGAAT TTCTTTTTAT TTGCACCCTC 120

ATCAGGATTA CAGCTTTATC AGGACTGCAT CTTCTTCAGA AATGAATATT TCTCTTACAA 180

CGCAAAGAAA GAAAAATCAA AATAAATTTT CTGATTGAAA ATGTAAAAAG GCAAATATTT 240

TTACAGTTTT AACTTTAATT TTTTATTGAG GACCAACTGT TTGAAAAATT CTCATTAGTC 300

ATTCCTTTAA ATTATGTGTA TGTGAGAGAA AGACGTAAGA TGGTTAATTA TTTCAAATGA 360

TGCAGTATAA AGAAGGGGCA TTATCACGGC AGAAACGAAA AAAGATATTT GTAGCTGGAG 420

GTTTTTATAG TCTAACATAT GGTTGCTATT TGTTCTACAA ATCCTTTTGA ATAATTTAAT 480

ATAGAGATTT CGAATAGAAA ATAATACTTT AGATAGAAAT TAATGAGTTT ATTATAACCA 540

TATATTATAA TAATTTACTT AGGAATTCTC TTTGATAAGA AACAAATGAA CTGAATGCAA 600

TTTTCTCCAC AGACCATATA AGACTGCCTA TGTACCTCCT CCTACATGCC ATTGGTTAAC 660

CATCAGTCAG TTTGCAGGGG TGGGGGGAGG GGTTTCCTGC CCATTGTTTT TGTAATCTCT 720

GAGGAGAAGC 730

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 121 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 38..121

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

AGCAGCAAAC ATTTGCTAGT CAGACAAGTG ACAGGGAATG GATTCCAAAC AGCAGTGTGT 60

AAAGCTAAAT GATGGCCACT TCATGCCTGT ATTGGGATTT GGCACCTATG CACCTCCAGA 120

G 121

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 252 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GTAAGAATAA TTCCTTTTAG TTTTCGGATT TCAAAAGAAT AAACCTAGTA GAAGTGAAAC 60

CCGTATTGGG TTGTAAGGTT CGTGTTCCTA CCTTACTCTG GATGACTCAC TGGTCTAGGT 120

TTCCTAGGCT AGGAGAAAAA AGTAGGCAAT CCTTGTTCTG CATTGAGGTC CATTCCTATG 180

GTCACGTACT GCTTATTTTT CGTTTGTGCA CTGTTTCTTT CTTCTGTTCA TGTCTAGTTC 240

CCAGCTTGGC AG 252

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 410 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GGAAGTCTGA GTGAGCATTC TGTGTAATAT CACTGGGAGA GAACTCATAT GAGCTTGCAC 60

CGTTTCCCTT CTATACTCCA TGTGATTTTT ACCATGTATA ATATCACTAT ATTAAAAATA 120

ATTAGGACTA TTTCAGTCAT GTTAACTTTT CCAACAAATC ACTGAATCTG AGGGTGTTAT 180

GTGGTACCTC CATAACAGTG ATCAACCAGA GATTGCCTGA GACTGAAGGT GTTTCTGGGA 240

TGCTCAACCT TTATTACTAA CCAGGAAAGA CTCAGGCAAA CTGAGATGGA CTTTTCACCC 300

CACATACAGA CAGGAGGAAA AGCTGATTCT TGTATAAAAG TCAATGCTTG TGCCTGAACT 360

ACCTCTCAGC CACAGTGATC ACCAGATACT ACCTTTGGTT GCTCCTCCAG 410

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 168 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..168

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GTTCCGAGAA GTAAAGCTTT GGAGGTCACA AAATTAGCAA TAGAAGCTGG GTTCCGCCAT 60

ATAGATTCTG CTCATTTATA CAATAATGAG GAGCAGGTTG GACTGGCCAT CCGAAGCAAG 120

ATTGCAGATG GCAGTGTGAA GAGAGAAGAC ATATTCTACA CTTCAAAG 168

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 700 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GTACTGTGTC TATGATGAGC TTGTGTGCAC ATGTATTTAT TGTGATTGTG TGGAGATGAC 60
AATTCTATGA CTGGATGAGT AGTTGTGGGT GAATTTTGCT TCTGGGTTCA AATTTATTCA 120
CACATACTCA CATACTAAAA CTGAAATCAA AATCAAGGAA TGATGATCAC TTTTCATTTT 180
GGCTGTGTTC CAATTTATGA CCTGAAAGTC CCTTTACTTT TTTGAGCTTC AGCCGAGATC 240
AGTGTGATTT GACATGTGCT ATAGAATCAC AGAGAACAAT AATCATGTTA TGGTTTTTCT 300
TATCGCCTGG GTGATTTTCT AAGATTTCTT ATTATTCTCT CAATTGCTAT CTTTATCAGT 360
GAGATAGAAA GCAATATAAG AAAGCTCTGG GAGTATTAAA TAATAGACAC TTAAATTGTC 420
CTAAATTGTG TCCAGCATAG TGAGCATGTT CAAAACTTGT TTTACCCCCC TTTTATGTTG 480
CTTTAGTTTC TAAGCAACAT AAATAGCTAT TCTTAAGCAT TGGGTTGAAT GGATAGAAGA 540
ATTAGACTGT TAAAATGAGT TGTAAACTCT ACTGAAGATA ATTCAGGTAA CATCATAGTT 600
ATTACTTAAT ACTAATCTTT ACATTTTAAG AATTTACTCC TATCATTCAG TAGATGTACA 660
AACTATACAT CCAACGTATA ATAAAGTTTA TAAGGATAGG 700

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 73 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

ACTAGATGGC ACAAAGTAAT AAGATTTGCT CAAGCATTCA TTCAAAATCA CCTCCATTCT 60
TTAACCTCTG CAG 73

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 117 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..117

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CTTTGGTCCA CTTTTCATCG ACCAGAGTTG GTCCGACCAG CCTTGGAAAA CTCACTGAAA 60

AAAGCTCAAT TGGACTATGT TGACCTCTAT CTTATTCATT CTCCAATGTC TCTAAAG 117

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 152 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GTATGCAGTT TGTATGAGCA TAAAATTGCG CTTCTGCTGT CATTATAAAC ATTGTTTATC 60

TGGATAGTTG AACAGAGCTT TTTATTAGGA GGATGTAGGG ATTATCACAC AGAAGAAGAA 120

CCGTAAGTGG AACACCTAAT TTCCTTTCTT TC 152

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 208 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

ATATAATATT TGTAAGAGAT TAGAGGAAGC CTGTCTCCTG AATACATTCC TTATACCTTC 60

ATATGTAAAA CACTTAGCAC ATATCACTTT CTGGAGCATT GTACCACCTG TCTCATGGAG 120

GATTAGTGTC CTTAAAGGTA CCTGGGGTTA CAGCTATGAG TGGAGAAATT AATTTGTGAC 180

ATCATTAAAA TGACTGCTTC TATTTCAG 208

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 78 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..78

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CCAGGTGAGG AACTTTCACC AACAGATGAA AATGGAAAAG TAATATTTGA CATAGTGGAT 60

CTCTGTACCA CCTGGGAG 78

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 98 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GTGAGTGCTT GGCGGAGAGG ACACAGAGAA GGATGACAAA AAGAGAAAAT CTGTTTCCCA 60

GGTTCGATAG GAAAGAATGG AATATGCACC ATTAGATC 98

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 249 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GACAGGAATC TCTTTCCTTG CTTGTGCATT AATCTATGCA GTTTCCTAAG GAAGAGATAG 60

AAATTCTTAC TCTTGCTGCC TCTATCTTCT TCCCCTATTT GCTGTTTGAA TTTTTCTTTT 120

TTTGACAATC ACTGCTAGCT ATTTTCATTG TCATACTTTG AAAGTTGTTG CTCTCACAGT 180

TCTGTCTTGC ATTTACCGTG ATTTGCAGCC AACTGCACAA ATAATTCCTC ACAACCCCTT 240

TCTCCACAG 249

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 123 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..123

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GCCATGGAGA AGTGTAAGGA TGCAGGATTG GCCAAGTCCA TTGGGGTGTC AAACTTCAAC 60

CGCAGGCAGC TGGAGATGAT CCTCAACAAG CCAGGACTCA AGTACAAGCC TGTCTGCAAC 120

CAG 123

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 138 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GTGAGCTCCC TTGGCCTTCT CTCCTTTCGG TTCTTCATGC CCCCTCTTCC TGTCCTATTG 60

CCAAATATCT GTTTGTTTTG TCCCAGTTAT CTTTGTGAAG TAGAAGATTA TCTAGAGAGC 120

AAAGCTTCTG TCAAGAAA 138

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 189 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

ATTTCCATTT ATACTTTTAG AAGATATATA AAATTTATTT CTATGAAAAA GGTTATTACT 60

TGACAATAAT ATCCTCAGCT CAAATATAAT GCTATACTGA TTATTATTCA GCTTCCTTAC 120

TTTCATCTTT TCAATATTAA CATAACTATT TCATATAAAT TGATGCTTCT CTCTTTTGGT 180

CAACTGCAG 189

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 110 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..110

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GTAGAATGTC ATCCGTATTT CAACCGGAGT AAATTGCTAG ATTTCTGCAA GTCGAAAGAT 60

ATTGTTCTGG TTGCCTATAG TGCTCTGGGA TCTCAACGAG ACAAACGATG 110

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 136 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GTAATAAAAA CAATGGGACC TTTACATAAA CCTTCATTTT GCAGAAAATT TTTTAGTCAG 60

AGCATCCTCA GTTTCCTGTA GTTAAGTTTC AAGTGGCTCA TGGAGAGGAA AGAGAATTGC 120

GTTTCTGACG AGATCT 136

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 66 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

TTTAGGGAGC TGCCTAACAA ACTATCGGCA GCCTCAGGGC CTCAGCCTTT CTGCCTTTCC 60

TTCCAG 66

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 166 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..166

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GGTGGACCCG AACTCCCCGG TGCTCTTGGA GGACCCAGTC CTTTGTGCCT TGGCAAAAAA 60

GCACAAGCGA ACCCCAGCCC TGATTGCCCT GCGCTACCAG CTGCAGCGTG GGGTTGTGGT 120

CCTGGCCAAG AGCTACAATG AGCAGCGCAT CAGACAGAAC GTGCAG 166

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 136 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GTGAGGAGCG GGGCTGTGGG CCTCAGGTCT CCTGCACAGT GTCCTTCACA CGTGTGCTTC 60

CTGTAAGGCT CTCAGGACAG CCTTGGGCCA GCTCCATTTC CCTGTATTTC CCATATGAAT 120

GCTTTGCGTG CATCCT 136

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 286 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CCCTATCATG TGGGCACAAT GTCAGCGCTG TTTCTTCTCC ATTTTCTGTT GAAATTTTCT 60

CTTTGTCTGC AGAGTTGCAC AGTTTCAATA CATAATATCT AGGAATGGAT TTCTGCTTAT 120

CTTTCGTGAG CTATTCATTG ACCCACCTGA GTGTTTAGAG CTGACTTCTA TAACTGTTTA 180

AAACTTACCA ATATTTTAAG TATTGTCTCT GCACCCTACT GTCTAATATA CTTGGGGATT 240

CACAACTGGC AATCTAAAAA TAATAAAAGT TTTTTATTTC TGATAG 286

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 83 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..83

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GTTTTTGAGT TCCAGTTGAC TGCAGAGGAC ATGAAAGCCA TAGATGGCCT AGACAGAAAT 60

CTCCACTATT TTAACAGTGA TAG 83

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 713 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GTAAGTTTCC TTTGTAAATG GGTGATCTAA TTTATTTCTG GAGAAGGAAT GTAGGATGGG 60

TGTTGAGAGT GACCTCCATA CCAGAGGGAC AGAGGCCAAT GTGAGTCAGA GGTGAGACTG 120

GAACTCTCCT GCTGGATTCA CTCCAGAGCT CTGTTCTCTG GCAGGGTGAG TGGGCAGGGA 180

TCAGCATGGG TCAACCTGTG CCTCTGCTCT CCTGACTCCA TGGAACTTTC CAGAGCAGCC 240

AACATCATTG CCAAGTCTGC ACGTTCCATA TAGGCCTGGT GTTTCTACCA CTGGACATGC 300

TGTGGATACT GCCCATGTGA CTTCATTAGA TGTTTCCAAA TCTGTGCTTA TATCACATTG 360

TCCCAAACCT GCTCAGCTCC TTATCAAATC AAAAACATTT CCATCAACTT TGTGGTCCAG 420

GTGCCAATTC CCACCTCCTT CATATGGAAT TGCTTGCTAG ATCCTGTCAA TTCAGCATCT 480

TTTATTATTT CAAATGTTTT TCCTCCTTCT CCTTGCACGT TTGTTCATGC CCCAAACTCT 540

GCTTTTGCCT CCAGAAAGCC TTCCTTAGTG GAGTGAATAG GAGTGCTTGT CCTTGATTTC 600

CTGCAATATG GAGCTCTCAA GGCAGAGAAT TTAAAAAAAT TTAAAATCAA GGAGTGTGAG 660

TGTGGAGGCA GAAGCTCCAT TGTTGTATAT AATTTGTAGC TGATAAAAGA TCT 713

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 415 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

TTTAATGCAC TGTAGCTCCT TGGATATTAG ACCCTATATC ATATATAACA ATTTACATTT 60

CTGAATCTTA CAAAATATAT TGCATACAGT AGGCAGTAGC AGGTAATAAG TAAAGTAACA 120

AAAGAAAGTA TAATCAGAGT ATCTCTGCTC TGCTGACAGA TGTACAGGAA TATACTTGAA 180

TATTTGACTT TGTGTGTTTT ACGTGTTAAC TTCCAGATAA GGGAATATGA TTGAATAATT 240

TATTATTTTG AAAATACTGT ATTATGAAGC CATGTTCATA AAGGTAAGAA AGGCAGATTC 300

TACAACTAGT CAGACAACTT AACATTCATA CTAATGACAG CTTCATTGAA ATCACTTTAC 360

TACTCCCCTA GTAATGGAGT CATTGCATTT ATATTATACA TTATTCTCTT TTCAG 415

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 230 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..230

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

TTTTGCTAGC CACCCTAATT ATCCATATTC AGATGAATAT TAACATGGAG GGCTTTGCCT 60

GATGATGTCT ACCAGAAGGC CCTGTGTGTG GATGGTGACG CAGAGGACGT CTCTATGCCG 120

GTGACTGGAC ATATCACCTC TACTTAAATC CGTCCTGTTT AGCGACTTCA GTCAACTACA 180

GCTGAGTCCA TAGGCCAGAA AGACAATAAA TTTTTATCAT TTTGAAATAA 230

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 109 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

TTGAATGTTT TCTCAAAGAT TCTTTACCTA CTCTGTTCTG TAGTGTGTGT TTTCTTCTGG 60

CTCAGAAGTG TGTGTGTGTG TGTGTGTGCT TTCTTCTGGC TCAACAGGG 109

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

TTTAGCTTTA CACACTGCTG TT 22

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

TCCAAAGCTT TACTTCTCGG 20

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

GATGAAAAGT GGACCA 16

2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

ATCTGTTGGT GAAAGTTC 18

2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

TCCAGCTGCC TGCGGT 16

2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

CTTGTACTTG AGTCCTG 17

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

CTCCGGTTGA AATACGGA 18

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

CATCGTTTGT CTCGTTGAGA 20

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

TCACTGTTAA AATAGTGGAG AT 22

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

ATCTGAATAT GGATAAT 17

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

TTCTCGGAAC CTGGAGGAGC 20

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GACACAGTAC CTTTGAAGTG 20

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

TGGACCAAAG CTGCAGAGGT 20

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

CCTCACCTGG CTGAAATAGA 20

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

AAGCACTCAC CTCCCAGGTG 20

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

GACATTCTAC CTGCAGTTGA 20

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

CTCAAAAACC TATCAGAAA 19

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

GGAAACTTAC CTATCACTGT 20

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

GCTAGCAAAA CTGAAAAGAG 20

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

GAGAAATATT CATTCTG 17

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

CGAGTCCTGA TAAAGCTG 18

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

GATGAGGGTG CAAATAA 17

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

GGAGTGTTAA TTAATAACAG TTT 23

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

CAGAGATTAC AAAAACAAT 19

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

TGCCTTTTTA CATTTTCAAT CA 22

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

ACACATAATT TAAAGGA 17

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

TTAAATTATT CAAAAGG 17

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

AAGAGAAATA TTCATTTCTG 20

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

CCCCTCCCCC CACCCCTGCA 20

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

CTGCCGTGAT AATGCCCC 18

CLAIMS

We claim:

1. An isolated nucleotide sequence encoding type 5 17β-hydroxysteroid dehydrogenase, said sequence being sufficiently homologous to SEQ ID No. 1 or a complement thereof, to hybridize under stringent conditions to the coding region of SEQ ID No. 1 or a complement thereof and said sequence encoding an enzyme which catalyzes the conversion of progesterone to 20∝-hydroxyprogesterone and the conversion of 4-androstenedione to testosterone.

2. The nucleotide sequence, as recited in claim 1, wherein said sequence is the coding region of SEQ ID No. 1.

3. A recombinant expression vector comprising a promoter sequence and a nucleotide sequence in accordance with claim 1.

4. A recombinant expression vector comprising a promoter sequence and a nucleotide sequence in accordance with claim 2.

5. A recombinant host cell, transformed or transfected with the vector of claim 4.

6. The recombinant host cell of claim 5, wherein said host cell is a eukaryotic cell.

7. A recombinant host cell, transformed or transfected with the vector of claim 3.

8. The recombinant host cell of claim 7, wherein said host cell is a eukaryotic cell.

9. The recombinant host cell of claim 8, wherein a nucleotide sequence that hybridizes under stringent conditions with SEQ ID No. 1 or its complement is integrated into the genome of said host cell.

10. The recombinant host cell of claim 9, wherein said nucleotide sequence is located on a recombinant vector.

11. The recombinant host cell, as recited in claim 8, wherein said host cell is capable of expressing a biologically active type 5 17β-hydroxysteroid dehydrogenase.

12. An isolated nucleotide sequence comprising at least ten consecutive nucleotides identical to 10 consecutive nucleotides in the coding region of SEQ ID No. 1, or the complement thereof.

13. The nucleotide sequence, as recited in claim 12, wherein said sequence comprises at least fifteen consecutive nucleotides identical to 15 consecutive nucleotides in the coding region of SEQ ID No. 1, or the complement thereof.

14. The nucleotide sequence, as recited in claim 13, wherein said sequence comprises at least twenty consecutive nucleotides identical to 20 consecutive nucleotides in the coding region of SEQ ID No. 1, or the complement thereof.

15. The nucleotide sequence, as recited in claim 13, wherein said sequence comprises at least thirty consecutive nucleotides identical to 30 consecutive nucleotides in the coding region of SEQ ID No. 1, or the complement thereof.

16. An oligonucleotide sequence selected from the group consisting of TTTAGCTTTACACACTGCTGTT (SEQ ID No. 30), TCCAAAGCTTTACTTCTCGG (SEQ ID No. 31), GATGAAAAGTGGACCA (SEQ ID No. 32), ATCTGTTGGTGAAAGTTC (SEQ ID No. 33), TCCAGCTGCCTGCGGT (SEQ ID No. 34), CTTGTACTTGAGTCCTG (SEQ ID No. 35), CTCCGGTTGAAATACGGA (SEQ ID No. 36), CATCGTTTGTCTCGTTGAGA (SEQ ID No. 37), TCACTGTTAAAATAGTGGAGAT (SEQ ID No. 38), and ATCTGAATATGGATAAT (SEQ ID No. 39).

17. An oligonucleotide sequence selected from the group consisting of TTCTCGGAACCTGGAGGAGC (SEQ ID No. 40), GACACAGTACCTTTGAAGTG (SEQ ID No. 41), TGGACCAAAGCTGCAGAGGT (SEQ ID No. 42), CCTCACCTGGCTGAAATAGA (SEQ ID No. 43), AAGCACTCACCTCCCAGGTG (SEQ ID No. 44), GACATTCTACCTGCAGTTGA (SEQ ID No. 45), CTCAAAAACCTATCAGAAA (SEQ ID No. 46), GGAAACTTACCTATCACTGT (SEQ ID No. 47), and GCTAGCAAAACTGAAAAGAG (SEQ ID No. 48).

18. An oligonucleotide sequence selected from the group consisting of GAGAAATATTCATTCTG (SEQ ID No. 49), CGAGTCCTGATAAAGCTG (SEQ ID No. 50), GATGAGGGTGCAAATAA (SEQ ID No. 51), GGAGTGTTAATTAATAACAGTTT (SEQ ID No. 52), CAGAGATTACAAAAACAAT (SEQ ID No. 53), TGCCTTTTTACATTTTCAATCA (SEQ ID No. 54), ACACATAATTTAAAGGA (SEQ ID No. 55), TTAAATTATTCAAAAGG (SEQ ID No. 56), AAGAGAAATATTCATTTCTG (SEQ ID No. 57), CCCCTCCCCCCACCCCTGCA (SEQ ID No. 58), and CTGCCGTGATAATGCCCC (SEQ ID No. 59).

19. A recombinant expression vector comprising:

a promoter sequence; and

an oligonucleotide sequence selected from the group consisting of TTTAGCTTTACACACTGCTGTT (SEQ ID No. 30), TCCAAAGCTTTACTTCTCGG (SEQ ID No. 31), GATGAAAAGTGGACCA (SEQ ID No. 32), ATCTGTTGGTGAAAGTTC (SEQ ID No. 33), TCCAGCTGCCTGCGGT (SEQ ID No. 34), CTTGTACTTGAGTCCTG (SEQ ID No. 35), CTCCGGTTGAAATACGGA (SEQ ID No. 36), CATCGTTTGTCTCGTTGAGA (SEQ ID No. 37), TCACTGTTAAAATAGTGGAGAT (SEQ ID No. 38), and ATCTGAATATGGATAAT (SEQ ID No. 39).

20. A recombinant expression vector comprising:

a promoter sequence; and

an oligonucleotide sequence selected from the group consisting of TTCTCGGAACCTGGAGGAGC (SEQ ID No. 40), GACACAGTACCTTTGAAGTG (SEQ ID No. 41), TGGACCAAAGCTGCAGAGGT (SEQ ID No. 42), CCTCACCTGGCTGAAATAGA (SEQ ID No. 43), AAGCACTCACCTCCCAGGTG (SEQ ID No. 44), GACATTCTACCTGCAGTTGA (SEQ ID No. 45), CTCAAAAACCTATCAGAAA (SEQ ID No. 46), GGAAACTTACCTATCACTGT (SEQ ID No. 47), and GCTAGCAAAACTGAAAAGAG (SEQ ID No. 48).

21. A recombinant expression vector comprising:

a promoter sequence; and

an oligonucleotide sequence selected from the group consisting of GAGAAATATTCATTCTG (SEQ ID No. 49), CGAGTCCTGATAAAGCTG (SEQ ID No. 50), GATGAGGGTGCAAATAA (SEQ ID No. 51), GGAGTGTTAATTAATAACAGTTT (SEQ ID No. 52), CAGAGATTACAAAAACAAT (SEQ ID No. 53), TGCCTTTTTACATTTTCAATCA (SEQ ID No. 54), ACACATAATTTAAAGGA (SEQ ID No. 55), TTAAATTATTCAAAAGG (SEQ ID No. 56), AAGAGAAATATTCATTTCTG (SEQ ID No. 57), CCCCTCCCCCCACCCCTGCA (SEQ ID No. 58), and CTGCCGTGATAATGCCCC (SEQ ID No. 59).

22. A method of blocking synthesis of type 5 17β-HSD, comprising the step of:

introducing an oligonucleotide selected from the group consisting of TTTAGCTTTACACACTGCTGTT (SEQ ID No. 30), TCCAAAGCTTTACTTCTCGG (SEQ ID No. 31), GATGAAAAGTGGACCA (SEQ ID No. 32), ATCTGTTGGTGAAAGTTC (SEQ ID No. 33), TCCAGCTGCCTGCGGT (SEQ ID No. 34), CTTGTACTTGAGTCCTG (SEQ ID No. 35), CTCCGGTTGAAATACGGA (SEQ ID No. 36), CATCGTTTGTCTCGTTGAGA (SEQ ID No. 37),

TCACTGTTAAAATAGTGGAGAT (SEQ ID No. 38), and ATCTGAATATGGATAAT (SEQ ID No. 39) into cells.

23. A method of blocking synthesis of type 5 17β-HSD, comprising the step of:

introducing an oligonucleotide selected from the group consisting of TTCTCGGAACCTGGAGGAGC (SEQ ID No. 40), GACACAGTACCTTTGAAGTG (SEQ ID No. 41), TGGACCAAAGCTGCAGAGGT (SEQ ID No. 42), CCTCACCTGGCTGAAATAGA (SEQ ID No. 43), AAGCACTCACCTCCCAGGTG (SEQ ID No. 44), GACATTCTACCTGCAGTTGA (SEQ ID No. 45), CTCAAAAACCTATCAGAAA (SEQ ID No. 46), GGAAACTTACCTATCACTGT (SEQ ID No. 47), and GCTAGCAAAACTGAAAAGAG (SEQ ID No. 48) into cells.

24. A method of blocking synthesis of type 5 17β-HSD, comprising the step of:

introducing an oligonucleotide selected from the group consisting of GAGAAATATTCATTCTG (SEQ ID No. 49), CGAGTCCTGATAAAGCTG (SEQ ID No. 50), GATGAGGGTGCAAATAA (SEQ ID No. 51), GGAGTGTTAATTAATAACAGTTT (SEQ ID No. 52), CAGAGATTACAAAAACAAT (SEQ ID No. 53), TGCCTTTTTACATTTTCAATCA (SEQ ID No. 54), ACACATAATTTAAAGGA (SEQ ID No. 55), TTAAATTATTCAAAAGG (SEQ ID No. 56), AAGAGAAATATTCATTTCTG (SEQ ID No. 57), CCCCTCCCCCCACCCCTGCA (SEQ ID No. 58), and CTGCCGTGATAATGCCCC (SEQ ID No. 59) into cells.

25. An isolated chromosomal DNA fragment which upon transcription and translation encodes type 5 17β-hydróxysteroid dehydrogenase and wherein said fragment contains nine exons and wherein said fragment includes introns which are 16 kilobase pairs in length.

26. An isolated DNA sequence encoding type 5 17β-hydroxysteroid dehydrogenase, said sequence being sufficiently homologous to SEQ ID No. 3 or a

complement thereof, to hybridize under stringent conditions to SEQ ID No. 3, or its complement.

27. A method for producing type 5 17β-hydroxysteroid dehydrogenase, comprising the steps of:

preparing a recombinant host transformed or transfected with the vector of claim 3; and

culturing said host under conditions which are conducive to the production of type 5 17β-hydroxysteroid dehydrogenase by said host.

28. A method for determining the inhibitory effect of a test compound on the enzymatic activity of type 5 17β-hydroxysteroid dehydrogenase, comprising the steps of:

providing type 5 17β-hydroxysteroid dehydrogenase;

contacting said type 5 17β-hydroxysteroid dehydrogenase with said test compound; and thereafter

determining the enzymatic activity of said type 5 17β-hydroxysteroid dehydrogenase in the presence of said test compound.

29. The method, as recited claim 28, wherein said step of determining enzymatic activity includes the steps of:

adding a substrate which is metabolized by said type 5 17β-hydroxysteroid dehydrogenase; and

determining an amount of said substrate which is converted to metabolite.

30. A method of interfering with the expression of type 5 17β-hydroxysteroid dehydrogenase, comprising the step of administering nucleic acids substantially identical to at least 15 consecutive nucleotides of SEQ ID No. 1 or a complement thereof.

31. A method of interfering with the synthesis of type 5 17β-hydroxysteroid dehydrogenase, comprising the step of administering antisense RNA complementary to mRNA encoded by at least 15 consecutive nucleotides of SEQ ID No. 1 or a

complement thereof.

32. A method of interfering with the expression of type 5 17β-hydroxysteroid dehydrogenase, comprising the step of administering nucleic acids substantially identical to at least 15 consecutive nucleotides of SEQ ID No. 3 or a complement thereof.

33. A method of interfering with the synthesis of type 5 17β-hydroxysteroid dehydrogenase, comprising the step of administering antisense RNA complementary to mRNA encoded by at least 15 consecutive nucleotides of SEQ ID No. 3 or a complement thereof.

34. A method for determining the inhibitory effect of antisense nucleic acids on the enzymatic activity of type 5 17β-hydroxysteroid dehydrogenase, comprising the steps of:

providing a host system capable of expressing type 5 17β-hydroxysteroid dehydrogenase;

introducing said antisense nucleic acids into said host system; and thereafter

determining the enzymatic activity of said type 5 17β-hydroxysteroid dehydrogenase.

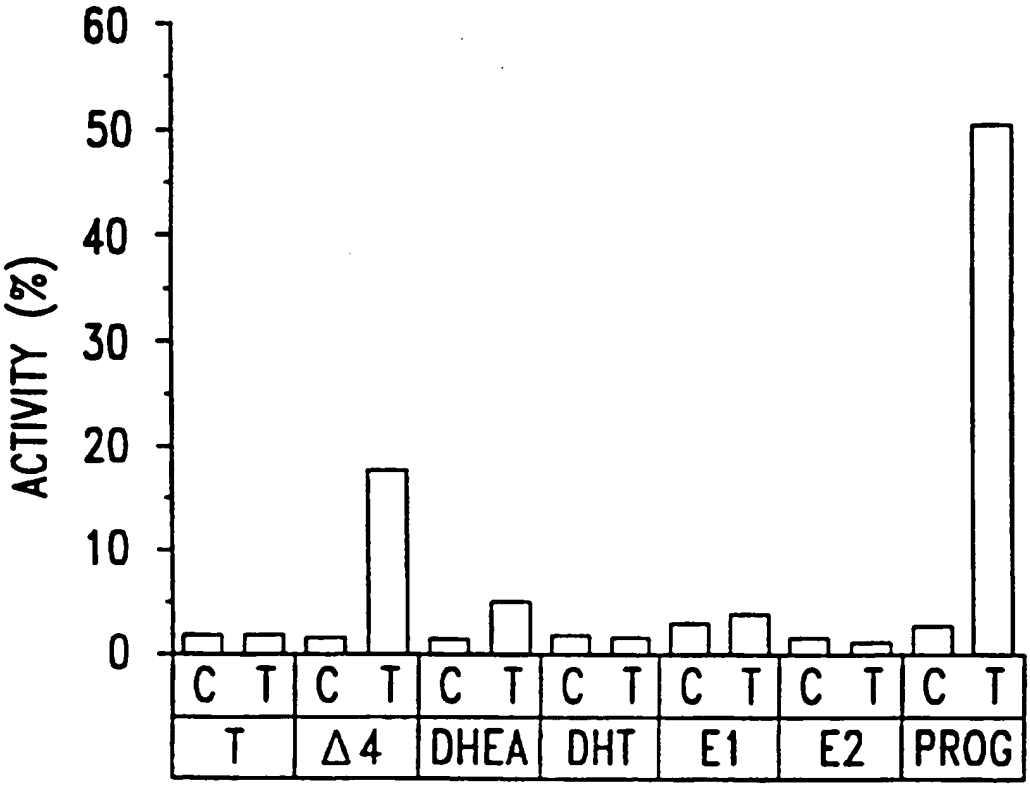
60
50
40
30
20
10
0
ACTIVITY (%)
C T C T C T C T C T C T C T
T Δ4 DHEA DHT E1 E2 PROG

FIG. 1A

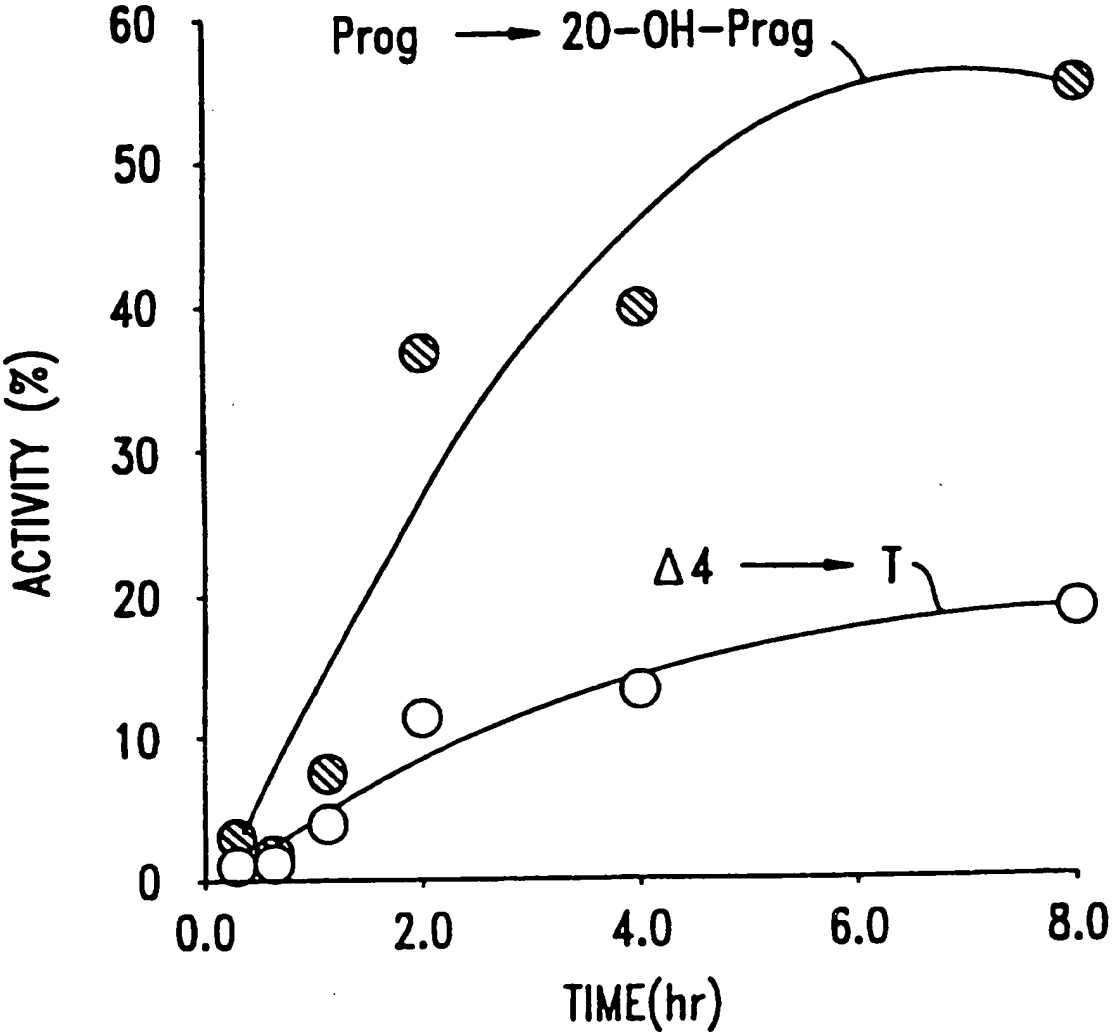
60
50
40
30
20
10
0
ACTIVITY (%)
Prog → 20-OH-Prog
Δ4 → T
0.0
2.0
4.0
6.0
8.0
TIME(hr)

FIG. 1B

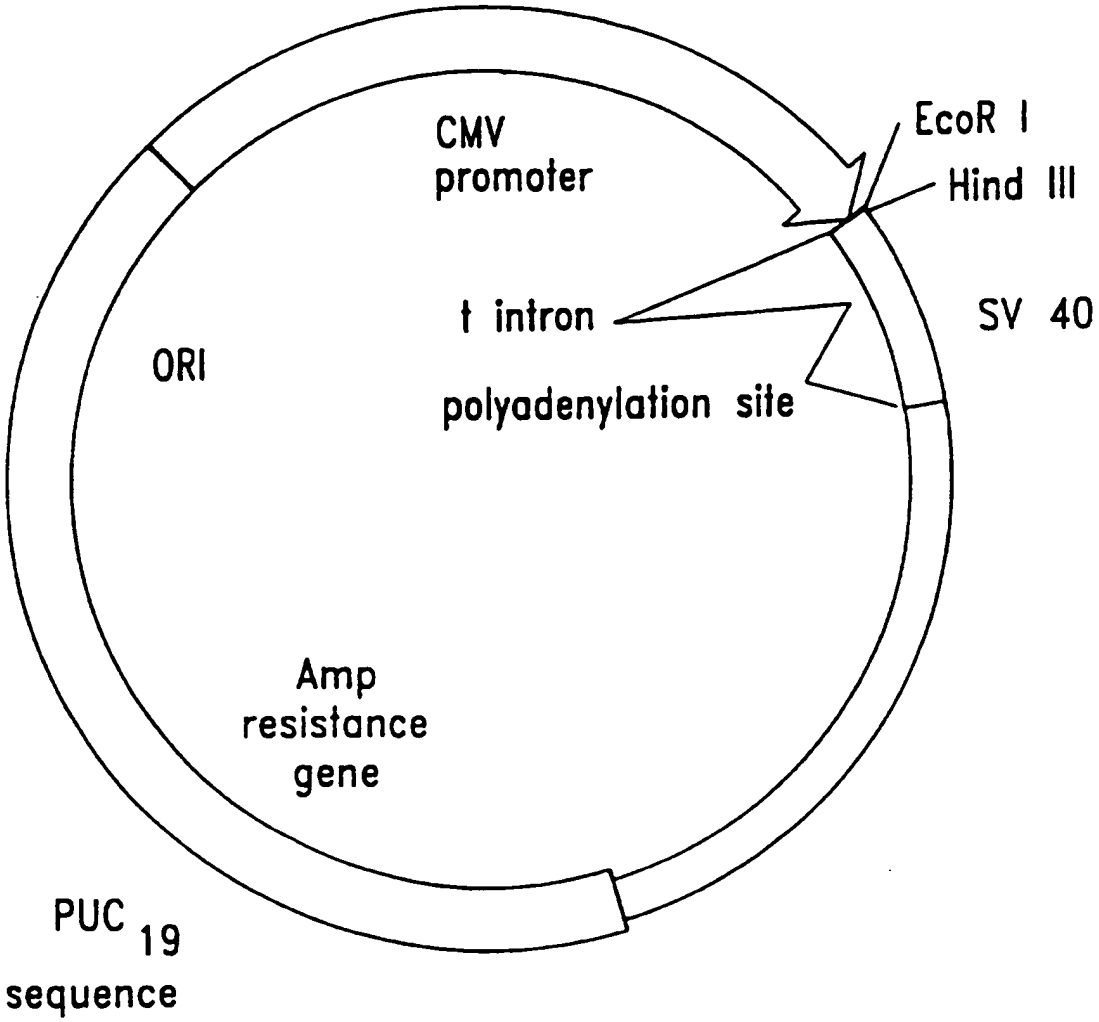
CMV
promoter
EcoR I
Hind III
t intron
SV 40
ORI
polyadenylation site
Amp
resistance
gene
PUC 19
sequence

FIG. 2

```
GTGACAGGGA ATG GAT TCC AAA CAG CAG TGT GTA AAG CTA AAT GAT GGC      49
           Met Asp Ser Lys Gln Gln Cys Val Lys Leu Asn Asp Gly

CAC TTC ATG CCT GTA TTG GGA TTT GGC ACC TAT GCA CCT CCA GAG GTT     97
His Phe Met Pro Val Leu Gly Phe Gly Thr Tyr Ala Pro Pro Glu Val

CCG AGA AGT AAA GCT TTG GAG GTC ACC AAA TTA GCA ATA GAA GCT GGG    145
Pro Arg Ser Lys Ala Leu Glu Val Thr Lys Leu Ala Ile Glu Ala Gly

TTC CGC CAT ATA GAT TCT GCT CAT TTA TAC AAT AAT GAG GAG CAG GTT    193
Phe Arg His Ile Asp Ser Ala His Leu Tyr Asn Asn Glu Glu Gln Val
                                    TSK
GGA CTG GCC ATC CGA AGC AAG ATT GCA GAT GGC AGT GTG AAG AGA GAA    241
Gly Leu Ala Ile Arg Ser Lys Ile Ala Asp Gly Ser Val Lys Arg Glu
                                          PKC
GAC ATA TTC TAC ACT TCA AAG CTT TGG TCC ACT TTT CAT CGA CCA GAG    289
Asp Ile Phe Tyr Thr Ser Lys Leu Trp Ser Thr Phe His Arg Pro Glu
                PKC
TTG GTC CGA CCA GCC TTG GAA AAC TCA CTG AAA AAA GCT CAA TTG GAC    337
Leu Val Arg Pro Ala Leu Glu Asn Ser Leu Lys Lys Ala Gln Leu Asp
```

FIG. 3A-1

```
TAT GTT GAC CTC TAT CTT ATT CAT TCT CCA ATG TCT CTA AAG CCA GGT     385
Tyr Val Asp Leu Tyr Leu Ile His Ser Pro Met Ser Leu Lys Pro Gly
                                           PKC
GAG GAA CTT TCA CCA ACA GAT GAA AAT GGA AAA GTA ATA TTT GAC ATA     433
Glu Glu Leu Ser Pro Thr Asp Glu Asn Gly Lys Val Ile Phe Asp Ile
           CK2
GTG GAT CTC TGT ACC ACC TGG GAG GCC ATG GAG AAG TGT AAG GAT GCA     481
Val Asp Leu Cys Thr Thr Trp Glu Ala Met Glu Lys Cys Lys Asp Ala
               CK2
GGA TTG GCC AAG TCC ATT GGG GTG TCA AAC TTC AAC CGC AGG CAG CTG     529
Gly Leu Ala Lys Ser Ile Gly Val Ser Asn Phe Asn Arg Arg Gln Leu
NM Aldo/Keto reductase family signature 1
GAG ATG ATC CTC AAC AAG CCA GGA CTC AAG TAC AAG CCT GTC TGC AAC     577
Glu Met Ile Leu Asn Lys Pro Gly Leu Lys Tyr Lys Pro Val Cys Asn

CAG GTA GAA TGT CAT CCG TAT TTC AAC CGG AGT AAA TTG CTA GAT TTC     625
Gln Val Glu Cys His Pro Tyr Phe Asn Arg Ser Lys Leu Leu Asp Phe
                                   NG
TGC AAG TCG AAA GAT ATT GTT CTG GTT GCC TAT AGT GCT CTG GGA TCT     673
Cys Lys Ser Lys Asp Ile Val Leu Val Ala Tyr Ser Ala Leu Gly Ser
                                                        PKC
CAA CGA GAC AAA CGA TGG GTG GAC CCG AAC TCC CCG GTG CTC TTG GAG     721
Gln Arg Asp Lys Arg Trp Val Asp Pro Asn Ser Pro Val Leu Leu Glu
```

FIG. 3A-2

```
GAC CCA GTC CTT TGT GCC TTG GCA AAA AAG CAC AAG CGA ACC CCA GCC     769
Asp Pro Val Leu Cys Ala Leu Ala Lys Lys His Lys Arg Thr Pro Ala

CTG ATT GCC CTG CGC TAC CAG CTG CAG CGT GGG GTT GTG GTC CTG GCC     817
Leu Ile Ala Leu Arg Tyr Gln Leu Gln Arg Gly Val Val Val Leu Ala
                                  Aldo/Keto reductase family
AAG AGC TAC AAT GAG CAG CGC ATC AGA CAG AAC GTG CAG GTT TTT GAG     865
Lys Ser Tyr Asn Glu Gln Arg Ile Arg Gln Asn Val Gln Val Phe Glu
    CK2      signature 2
TTC CAG TTG ACT GCA GAG GAC ATG AAA GCC ATA GAT GGC CTA GAC AGA     913
Phe Gln Leu Thr Ala Glu Asp Met Lys Ala Ile Asp Gly Leu Asp Arg
                                                 NM
AAT CTC CAC TAT TTT AAC AGT GAT AGT TTT GCT AGC CAC CCT AAT TAT     961
Asn Leu His Tyr Phe Asn Ser Asp Ser Phe Ala Ser His Pro Asn Tyr

CCA TAT TCA GAT GAA TAT TAA CATGGAGACT TTGCCTGATG ATGTCTACCA      1012
Pro Tyr Ser Asp Glu Tyr  *

GAAGGCCCTG TGTGTGGATG GTGACGCAGA GGACGTCTCT ATGCCGGTGA CTGGACATAT 1072
CACCTCTACT TAAATCCGTC CTGTTTAGCG ACTTCAGTCA ACTACAGCTC ACTCCATAGG 1132
CCAGAAATAC AATAAATCCT GTTTAGCGAC TTCAGTCAAC TACAGCTCAC TCCATAGGCC 1192
AGAAATACAA TAAA                                                   1206
```

FIG. 3A-3

```
h 20αHSD    MDSKQQCVKLNDGHFMPVLGFGTYAPPEVPRSKALEVTKLAIEAGFRHIDSAHLY   55
rb20αHSD    --P-F-R-A-S-----I----------E---K---M-A--I--D----------YF-
r 20αHSD    -N--I-KME------SI----------TE-NL-K-SM-S--I--DV-------CS---
b 20αHSD    ............-AK--I--L--WKS-PGKVTE-VK-AIDLGY...----C--V-
h 3αHSD     --P-Y-R-E----------------------NR-V----------------Y--
r 3αhsd     ---ISLR-A----N-I-------TV-EK-AKDEVIKA--I--DN-----F---Y--
b pgfs      --P-S-R--------I----------E---K-E---A--F---V-----V------
f ρ-crys    TLT-ETR-T----NM--I--L----S-H--K-L-E-AV-I--DV-Y----C-FIT

h 20αHSD    NNEEQVGLAIRSKIADGSVKREDIFYTSKLWSTFHRPELVRPALENSLKKAQLDY  110
rb20αHSD    K--KE-----------T-------------C----------S--D---NL----
r 20αHSD    Q---EI-Q--V---E--T----------------S--------S------R-LN---
b 20αhsd    Q--NE----LQA-LQEKV------L-IV----C-Y-DKD--KG-CQKT-SDLK---
h 3αHSD     ------------------------------C--FQ-QM-Q----S----L----
r 3αHSD     EV--E--Q------E--T----------------------TC--KT--ST----
b pgfs      Q------Q---------T-------------CNSLQ---------K--QNL----
f ρ-crys    G--MHI-NG-----S--T---------G---C-YFS--M--KG--R--RDVGM--

h 20αHSD    VDLYLIHSPMSLKPGEELSPTDENGKVIFDIVDLCTTWEAMEKCKDAGLAKSIGV  165
rb20αHSD    ----I--F-TA----V-II----H--A---T--I-A----------------
r 20αHSD    -------F-V-----D--L-Q--H-NL-L-T----D----------------
b 20αhsd    L------W-TGF---KDFF-L--D-N--PSEK-FVD--T---ELV-E--V-A---
h 3αHSD     -----L-F--A-----TPL-K---------T---SA---V--------------
r 3αHSD     ----I--F--A-Q--DIFF-R--H--LL-ET--I-D----------------
b pgfs      ----I----V-----NKFV-K--S--L---S----H----L--------T-----
f ρ-crys    L--F-M-W-V----SGASD-S-KDKPF-Y-N----A----L-AR-----VR-L--
```

FIG. 4A-1

```
h 20αHSD    SNFNRRQLEMILNKPGLKYKPVCNQVECHPYFNRSKLLDFCKSKDIVLVAYSALG 220
rb20αHSD    ------------------------------L-QG---E-----G----------
r 20αHSD    ---------K--------HR---------L-L-Q----AY--MN-------G---
b 20αhsd    ----HL-V-K----------AV--I-----LTQE--IQY-N--G--VT---P--
h 3αHSD     ----C--------------------------L-Q----------------H----
r 3αHSD     ----C----R--------------------L-L-QG-M--Y------I--S-CT--
b pgfs      ----HK---K----------------------L-Q----E----H-------A---
f ρ-crys    ---------R--------------------V-L-QN--HSY---------T--V--

h 20αHSD    SQRDKRWVDPNSPVLLEDPVLCALAKKHKRTPALIALRYQLQRGVVVLAKSYNEQ 275
rb20αHSD    -H-EPE---QSA--------LIG-------QQ--------------I------FT-K
r 20αHSD    T--Y-YCINEDT----D--I--TM---YQ------------E--I-T-V--F--E
b 20αHSD    -.P-RP-AK-ED-SI----RIK-I-D-YNK-T-QVLI-FPI--NLI-IP--VTPE
h 3αHSD     T--HKL-----------------------------------------------
r 3αHSD     -S---T---QK-----D------I---Y-Q----V-----------P-IR-F-AK
b pgfs      A-LLSE--NS-N----------I-----Q----V-----V----------F-KK
f ρ-crys    -H--RN---LSL----D--I-NKV-A-YN--S-E--M-FI--KQI------FTPA

h 20αHSD    RIRQNVQVFEFQLTAEDMKAIDGLDRNLHYFNSDSFASHPNYPYSDEY       323
rb20αHSD    --KE-I--------PS----V--S-N--FR-VTA-FAIG-----F----
r 20αHSD    ---E-L---D---ASD--ELL-N-----R--PANM-KA---F-F----
b 20αHSD    --AE-F---D-E-DK---NT-LSYN-DWRACALV-C---RD--FHE-F
h 3αHSD     ---E----------S----VL---N--YR-VVM-FLMD--D--F----
r 3αHSD     --KELT--------AS-----L---N--FR-N-AKY-DD---H-FT--
b pgfs      --KE-M---D-E--P---------N--IR-YDFQKGIG--E--F-E--
f ρ-crys    --K--LG----E-KP----SLES-------GPFREVKQ--E--FH---
```

FIG. 4A-2

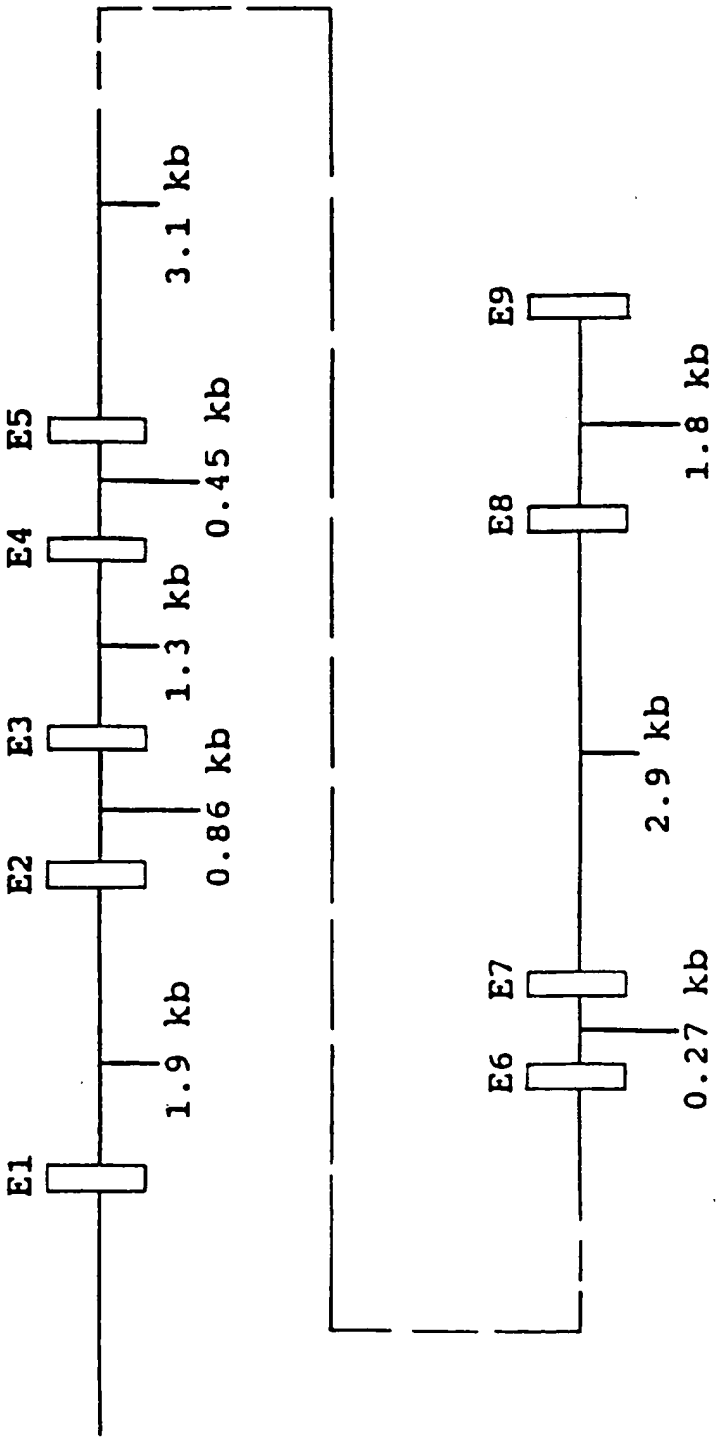
E1
E2
E3
E4
E5
E6
E7
E8
E9
1.9 kb
0.86 kb
1.3 kb
0.45 kb
3.1 kb
0.27 kb
2.9 kb
1.8 kb

FIG. 5

AAGAACAAATACTATTAAGGCACTGCTTGCATATATTAAATGATGTCCAAACTCCAAAAACTGTTAATAATTAACACTCC
AATAAAAACTACACCAGAATTTCTTTTTATTTGCACCCTCATCAGGATTACAGCTTTATCAGGACTGCATCTTCTTCAGA
AATGAATATTTCTCTTACAACGCAAAGAAAGAAAAATCAAAATAAATTTTCTGATTGAAAATGTAAAAAGGCAAATATTT
TTACAGTTTTAACTTTAATTTTTTATTGAGGACCAACTGTTTGAAAAATTCTCATTAGTCATTCCTTTAAATTATGTGTA
TGTGAGAGAAAGACGTAAGATGGTTAATTATTTCAAATGATGCAGTATAAAGAAGGGGCATTATCACGGCAGAAACGAAA
AAAGATATTTGTAGCTGGAGGTTTTTATAGTCTAACATATGGTTGCTATTTGTTCTACAAATCCTTTTGAATAATTTAAT
ATAGAGATTTCGAATAGAAAATAATACTTTAGATAGAAATTAATGAGTTTATTATAACCATATATTATAATAATTTACTT
AGGAATTCTCTTTGATAAGAAACAAATGAACTGAATGCAATTTTCTCCACAGACCATATAAGACTGCCTATGTACCTCCT
CCTACATGCCATTGGTTAACCATCAGTCAGTTTGCAGGGGTGGGGGGAGGGGTTTCCTGCCCATTGTTTTTGTAATCTCT
GAGGAGAAGC

AGCAGCAAACATTTGCTAGTCAGACAAGTGACAGGGA
Met Asp Ser Lys Gln Gln Cys Val Lys Leu Asn Asp Gly His Phe Met Pro Val Leu Gly 20
ATG GAT TCC AAA CAG CAG TGT GTA AAG CTA AAT GAT GGC CAC TTC ATG CCT GTA TTG GGA
Phe Gly Thr Tyr Ala Pro Pro Glu 28
TTT GGC ACC TAT GCA CCT CCA GAG

GTAAGAATAATTCCTTTTAGTTTTCGGATTTCAAAAGAATAAACCTAGTAGAAGTGAAACCCGTATTGGGTTGTAAGGTT
CGTGTTCCTACCTTACTCTGGATGACTCACTGGTCTAGGTTTCCTAGGCTAGGAGAAAAAAGTAGGCAATCCTTGTTCTG
CATTGAGGTCCATTCCTATGGTCACGTACTGCTTATTTTTCGTTTGTGCACTGTTTCTTTCTTCTGTTCATGTCTAGTTC
CCAGCTTGGCAG.....1.2 kb.....GGAAGTCTGAGTGAGCATTCTGTGTAATATCACTGGGAGAG
AACTCATATGAGCTTGCACCGTTTCCCTTCTATACTCCATGTGATTTTTACCATGTATAATATCACTATATTAAAAATAA
TTAGGACTATTTCAGTCATGTTAACTTTTCCAACAAATCACTGAATCTGAGGGTGTTATGTGGTACCTCCATAACAGTGA
TCAACCAGAGATTGCCTGAGACTGAAGGTGTTTCTGGGATGCTCAACCTTTATTACTAACCAGGAAAGACTCAGGCAAAC
TGAGATGGACTTTTCACCCCACATACAGACAGGAGGAAAAGCTGATTCTTGTATAAAAGTCAATGCTTGTGCCTGAACTA
CCTCTCAGCCACAGTGATCACCAGATACTACCTTTGGTTGCTCCTCCAG

FIG. 6A-1

```
Val Pro Arg Ser Lys Ala Leu Glu Val Thr Lys Leu Ala Ile Glu Ala Gly Phe Arg His  48
GTT CCG AGA AGT AAA GCT TTG GAG GTC ACA AAA TTA GCA ATA GAA GCT GGG TTC CGC CAT
Ile Asp Ser Ala His Leu Tyr Asn Asn Glu Glu Gln Val Gly Leu Ala Ile Arg Ser Lys  68
ATA GAT TCT GCT CAT TTA TAC AAT AAT GAG GAG CAG GTT GGA CTG GCC ATC CGA AGC AAG
Ile Ala Asp Gly Ser Val Lys Arg Glu Asp Ile Phe Tyr Thr Ser Lys                   84
ATT GCA GAT GGC AGT GTG AAG AGA GAA GAC ATA TTC TAC ACT TCA AAG

GTACTGTGTCTATGATGAGCTTGTGTGCACATGTATTTATTGTGATTGTGTGGAGATGACAATTCTATGACTGGATGAGT
AGTTGTGGGTGAATTTTGCTTCTGGGTTCAAATTTATTCACACATACTCACATACTAAAACTGAAATCAAAATCAAGGAA
TGATGATCACTTTTCATTTTGGCTGTGTTCCAATTTATGACCTGAAAGTCCCTTTACTTTTTTGAGCTTCAGCCGAGATC
AGTGTGATTTGACATGTGCTATAGAATCACAGAGAACAATAATCATGTTATGGTTTTTCTTATCGCCTGGGTGATTTTCT
AAGATTTCTTATTATTCTCTCAATTGCTATCTTTATCAGTGAGATAGAAAGCAATATAAGAAAGCTCTGGGAGTATTAAA
TAATAGACACTTAAATTGTCCTAAATTGTGTCCAGCATAGTGAGCATGTTCAAAACTTGTTTTACCCCCCTTTTATGTTG
CTTTAGTTTCTAAGCAACATAAATAGCTATTCTTAAGCATTGGGTTGAATGGATAGAAGAATTAGACTGTTAAAATGAGT
TGTAAACTCTACTGAAGATAATTCAGGTAACATCATAGTTATTACTTAATACTAATCTTTACATTTTAAGAATTTACTCC
TATCATTCAGTAGATGTACAAACTATACATCCAACGTATAATAAAGTTTATAAGGATAGG.....0.1 kb.....
ACTAGATGGCACAAAGTAATAAGATTTGCTCAAGCATTCATTCAAAATCACCTCCATTCTTTAACCTCTGCAG
```

FIG. 6A-2

```
Leu Trp Ser Thr Phe His Arg Pro Glu Leu Val Arg Pro Ala Leu Glu Asn Ser Leu Lys  104
CTT TGG TCC ACT TTT CAT CGA CCA GAG TTG GTC CGA CCA GCC TTG GAA AAC TCA CTG AAA
Lys Ala Gln Leu Asp Tyr Val Asp Leu Tyr Leu Ile His Ser Pro Met Ser Leu Lys      123
AAA GCT CAA TTG GAC TAT GTT GAC CTC TAT CTT ATT CAT TCT CCA ATG TCT CTA AAG

GTATGCAGTTTGTATGAGCATAAAATTGCGCTTCTGCTGTCATTATAAACATTGTTTATCTGGATAGTTGAACAGAGCTT
TTTATTAGGAGGATGTAGGGATTATCACACAGAAGAAGAACCGTAAGTGGAACACCTAATTTCCTTTCTTTC.........
0.9 kb..........ATATAATATTTGTAAGAGATTAGAGGAAGCCTGTCTCCTGAATACATTCCTTATACCTTCATAT
GTAAAACACTTAGCACATATCACTTTCTGGAGCATTGTACCACCTGTCTCATGGAGGATTAGTGTCCTTAAAGGTACCTG
GGGTTACAGCTATGAGTGGAGAAATTAATTTGTGACATCATTAAAATGACTGCTTCTATTTCAG

Pro Gly Glu Glu Leu Ser Pro Thr Asp Glu Asn Gly Lys Val Ile Phe Asp Ile Val Asp  143
CCA GGT GAG GAA CTT TCA CCA ACA GAT GAA AAT GGA AAA GTA ATA TTT GAC ATA GTG GAT
Leu Cys Thr Thr Trp Glu                          149
CTC TGT ACC ACC TGG GAG

GTGAGTGCTTGGCGGAGAGGACACAGAGAAGGATGACAAAAAGAGAAAATCTGTTTCCCAGGTTCGATAGGAAAGAATGG
AATATGCACCATTAGATC.......0.1 kb.............GACAGGAATCTCTTTCCTTGCTTGTGCATTAATCTAT
GCAGTTTCCTAAGGAAGAGATAGAAATTCTTACTCTTGCTGCCTCTATCTTCTTCCCCTATTTGCTGTTTGAATTTTTCT
TTTTTTGACAATCACTGCTAGCTATTTTCATTGTCATACTTTGAAAGTTGTTGCTCTCACAGTTCTGTCTTGCATTTACC
GTGATTTGCAGCCAACTGCACAAATAATTCCTCACAACCCCTTTCTCCACAG
```

FIG. 6A-3

```
Ala Met Glu Lys Cys Lys Asp Ala Gly Leu Ala Lys Ser Ile Gly Val Ser Asn Phe Asn  169
GCC ATG GAG AAG TGT AAG GAT GCA GGA TTG GCC AAG TCC ATT GGG GTG TCA AAC TTC AAC
Arg Arg Gln Leu Glu Met Ile Leu Asn Lys Pro Gly Leu Lys Tyr Lys Pro Val Cys Asn  189
CGC AGG CAG CTG GAG ATG ATC CTC AAC AAG CCA GGA CTC AAG TAC AAG CCT GTC TGC AAC
Gln                                                                              190
CAG

GTGAGCTCCCTTGGCCTTCTCTCCTTTCGGTTCTTCATGCCCCCTCTTCCTGTCCTATTGCCAAATATCTGTTTGTTTT
GTCCCAGTTATCTTTGTGAAGTAGAAGATTATCTAGAGAGCAAAGCTTCTGTCAAGAAA.............2.8 kb
.....ATTTCCATTTATACTTTTAGAAGATATATAAAATTTATTTCTATGAAAAAGGTTATT
ACTTGACAATAATATCCTCAGCTCAAATATAATGCTATACTGATTATTATTCAGCTTCCTTACTTTCATCTTTTCAATA
TTAACATAACTATTTCATATAAATTGATGCTTCTCTCTTTTGGTCAACTGCAG

Val Glu Cys His Pro Tyr Phe Asn Arg Ser Lys Leu Leu Asp Phe Cys Lys Ser Lys Asp  210
GTA GAA TGT CAT CCG TAT TTC AAC CGG AGT AAA TTG CTA GAT TTC TGC AAG TCG AAA GAT
Ile Val Leu Val Ala Tyr Ser Ala Leu Gly Ser Gln Arg Asp Lys Arg Tr                227
ATT GTT CTG GTT GCC TAT AGT GCT CTG GGA TCT CAA CGA GAC AAA CGA TG

GTAATAAAAACAATGGGACCTTTACATAAACCTTCATTTTGCAGAAAATTTTTTAGTCAGAGCATCCTCAGTTTCCTGT
AGTTAAGTTTCAAGTGGCTCATGGAGAGGAAAGAGAATTGCGTTTCTGACGAGATCT.............0.1 kb....
.....TTTAGGGAGCTGCCTAACAAACTATCGGCAGCCTCAGGGCCTCAGCCTTTCTGCCTTTCCTTCCAG
```

FIG. 6B-1

```
  p Val Asp Pro Asn Ser Pro Val Leu Leu Glu Asp Pro Val Leu Cys Ala Leu Ala Lys  246
  G GTG GAC CCG AAC TCC CCG GTG CTC TTG GAG GAC CCA GTC CTT TGT GCC TTG GCA AAA
Lys His Lys Arg Thr Pro Ala Leu Ile Ala Leu Arg Tyr Gln Leu Gln Arg Gly Val Val  266
AAG CAC AAG CGA ACC CCA GCC CTG ATT GCC CTG CGC TAC CAG CTG CAG CGT GGG GTT GTG
Val Leu Ala Lys Ser Tyr Asn Glu Gln Arg Ile Arg Gln Asn Val Gln                  282
GTC CTG GCC AAG AGC TAC AAT GAG CAG CGC ATC AGA CAG AAC GTG CAG

GTGAGGAGCGGGGCTGTGGGCCTCAGGTCTCCTGCACAGTGTCCTTCACACGTGTGCTTCTTGTAAGGCTCTCAGGACA
GCCTTGGGCCAGCTCCATTTCCCTGTATTTCCCATATGAATGCTTTGCGTGCATCCT.............2.5......
.........CCCTATCATGTGGGCACAATGTCAGCGCTGTTTCTTCTCCATTTTCTGTTGAAATTTTCTCTTTGTCTGC
AGAGTTGCACAGTTTCAATACATAATATCTAGGAATGGATTTCTGCTTATTTTTCGTGAGCTATTCATTGACCCACCTG
AGTGTTTAGAGCTGACTTCTATAACTGTTTAAAACTTACCAATATTTTAAGTATTGTCTCTGCACCCTACTGTCTAATA
TACTTGGGGATTCACAACTGGCAATCTAAAAATAATAAAAGTTTTTTATTTCTGATAG

Val Phe Glu Phe Gln Leu Thr Ala Glu Asp Met Lys Ala Ile Asp Gly Leu Asp Arg Asn  302
GTT TTT GAG TTC CAG TTG ACT GCA GAG GAC ATG AAA GCC ATA GAT GGC CTA GAC AGA AAT
Leu His Tyr Phe Asn Ser Asp Se                                                   310
CTC CAC TAT TTT AAC AGT GAT AG
```


FIG. 6B-2

```
GTAAGTTTCCTTTGTAAATGGGTGATCTAATTTATTTCTGGAGAAGGAATGTAGGATGGGTGTTGAGAGTGACCTCCATA
CCAGAGGGACAGAGGCCAATGTGAGTCAGAGGTGAGACTGGAACTCTCCTGCTGGATTCACTCCAGAGCTCTGTTCTCTG
GCAGGGTGAGTGGGCAGGGATCAGCATGGGTCAACCTGTGCCTCTGCTCTCCTGACTCCATGGAACTTTCCAGAGCAGCC
AACATCATTGCCAAGTCTGCACGTTCCATATAGGCCTGGTGTTTCTACCACTGGACATGCTGTGGATACTGCCCATGTGA
CTTCATTAGATGTTTCCAAATCTGTGCTTATATCACATTGTCCCAAACCTGCTCAGCTCCTTATCAAATCAAAAACATTT
CCATCAACTTTGTGGTCCAGGTGCCAATTCCCACCTCCTTCATATGGAATTGCTTGCTAGATCCTGTCAATTCAGCATCT
TTTATTATTTCAAATGTTTTTCCTCCTTCTCCTTGCACGTTTGTTCATGCCCCAAACTCTGCTTTTGCCTCCAGAAAGCC
TTCCTTAGTGGAGTGAATAGGAGTGCTTGTCCTTGATTTCCTGCAATATGGAGCTCTCAAGGCAGAGAATTTAAAAAAAT
TTAAAATCAAGGAGTGTGAGTGTGGAGGCAGAAGCTCCATTGTTGTATATAATTTGTAGCTGATAAAAGATCT.......
..2.7 kb.....TTTAATGCACTGTAGCTCCTTGGATATTAGACCCTATATCATATATAACAATTTACATTTCTG
AATCTTACAAAATATATTGCATACAGTAGGCAGTAGCAGGTAATAAGTAAAGTAACAAAAGAAAGTATAATCAGAGTATC
TCTGCTCTGCTGACAGATGTACAGGAATATACTTGAATATTTGACTTTGTGTGTTTTACGTGTTAACTTCCAGATAAGGG
AATATGATTGAATAATTTATTATTTTGAAAATACTGTATTATGAAGCCATGTTCATAAAGGTAAGAAAGGCAGATTCTAC
AACTAGTCAGACAACTTAACATTCATACTAATGACAGCTTCATTGAAATCACTTTACTACTCCCCTAGTAATGGAGTCAT
TGCATTTATATTATACATTATTCTCTTTTCAG

  r Phe Ala Ser His Pro Asn Tyr Pro Tyr Ser Asp Glu Tyr End                    323
  T TTT GCT AGC CAC CCT AAT TAT CCA TAT TCA GAT GAA TAT TAA

CATGGAGGGCTTTGCCTGATGATGTCTACCAGAAGGCCCTGTGTGTGGATGGTGACGCAGAGGACGTCTCTATGCCGGTG
ACTGGACATATCACCTCTACTTAAATCCGTCCTGTTTAGCGACTTCAGTCAACTACAGCTGAGTCCATAGGCCAGAAAGA
CAATAAATTTTTATCATTTTGAAATAA

TTGAATGTTTTCTCAAAGATTCTTTACCTACTCTGTTCTGTAGTGTGTGTTTTCTTCTGGCTCAGAAGTGTGTGTGTGTG
TGTGTGTGCTTTCTTCTGGCTCAACAGGG
```

FIG. 6B-3

INTERNATIONAL SEARCH REPORT

I. .national Application No
PCT/CA 96/00605

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N9/04

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CLINICAL AND INVESTIGATIVE MEDICINE, vol. 18, no. sup.4, September 1995, TORONTO CA, page b40 XP000196659 Y. ZHANG ET AL.: "Isolation and characterization of human type 5 17-beta-hydroxysteroid dehydrogenase"	1-15,25, 26
A	see the whole document	16-24, 27-34

X	EMBL SEQUENCE DATABASE, Acc.No.: Emhuml:Hsorf1,15 December 1993, N. Miyajima, "Human mRNA (HA1753)". see abstract XP002020808	1,2, 12-15
A		3-11, 16-24, 27-34

	-/--	

[X] Further documents are listed in the continuation of box C. [ ] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

1

Date of the actual completion of the international search	Date of mailing of the international search report
12 December 1996	06. 01. 97
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax (+ 31-70) 340-3016	Authorized officer De Kok, A

Form PCT ISA 210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JOURNAL OF STEROID BIOCHEMISTRY AND MOLECULAR BIOLOGY, vol. 46, 1993, OXFORD GB, pages 673-679, XP000196680 K.N. QIN ET AL.: "Molecular cloning of multiple cDNAs encoding human enzymes structurally related to 3-alpha-hydroxysteroid dehydrogenase"	1,2, 12-15
A	see the whole document	3-11, 16-24, 27-34

A	JOURNAL OF STEROID BIOCHEMISTRY AND MOLECULAR BIOLOGY, vol. 55, no. 5-6, 1995, OXFORD GB, pages 581-587, XP000196678 V. LUU-THE ET AL.: "Characterisitics of human types 1, 2 and 3 17-beta-hydroxysteroid dehydrogenase activities" see the whole document	3-11, 27-29

A	JOURNAL OF STEROID BIOCHEMISTRY, vol. 34, no. 1-6, 1989, OXFORD GB, pages 189-197, XP000196658 F. LABRIE ET AL.: "Characterization of two mRNA species encoding human estradiol 17 beta-dehydrogenase and assignment of the gene to chromosome 17" see the whole document	1-15

A	JOURNAL OF STEROID BIOCHEMISTRY AND MOLECULAR BIOLOGY, vol. 41, no. 3-8, March 1992, OXFORD GB, pages 605-608, XP000196681 M. DUMONT ET AL.: "Expression of human 17 beta-hydroxysteroid dehydrogenase in mammalian cells" see the whole document	1-11

A	EUROPEAN JOURNAL OF BIOCHEMISTRY, vol. 209, no. 1, 1992, BERLIN DE, pages 459-466, XP000196661 H. PELTOKETO ET AL.: "Genomic organisation and DNA sequences of human 17-beta hydroxysteroid dehydrogenase genes and flanking regions" see the whole document	1,25,26

	-/--	

1

INTERNATIONAL SEARCH REPORT

International Application No
PCT/CA 96/00605

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JOURNAL OF STEROID BIOCHEMISTRY AND MOLECULAR BIOLOGY, vol. 53, no. 1-6, 1995, OXFORD GB, pages 37-39, XP000196679 S. ANDERSSON ET AL.: "The molecular biology of androgenic 17-beta-hydroxysteroid dehydrogenases" see the whole document ---	1,25,26
A	JOURNAL OF THE NATIONAL CANCER INSTITUTE, vol. 81, no. 20, 18 October 1989, BETHESDA US, pages 39-44, XP000351436 M. ROTHENBERG ET AL.: "Oligodeoxynucleotides as anti-sense inhibitors of gene expression: therapeutic implications" see the whole document ---	16-24
P,X	EMBL SEQUENCE DATABASE, Acc. No.: Emhuml:Hs516761, 19 April 1996, P.J. Ciaccio et al.,"Human dihydrodiol dehydrogenase gene, 5'-flanking region" XP002020809 see abstract -----	26,32,33

1

Form PCT ISA 210 (continuation of second sheet) (July 1992)